



Essays on Macroeconomics with Financial Frictions

Dominik Thaler

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence, 07 October 2016

European University Institute
Department of Economics

Essays on Macroeconomics with Financial Frictions

Dominik Thaler

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Evi Pappa, EUI, Supervisor

Prof. Árpád Ábrahám, EUI

Prof. Luisa Lambertini, École Polytechnique Fédérale de Lausanne (EPFL)

Dr. Peter Karadi, European Central Bank

© Dominik Thaler, 2016

No part of this thesis may be copied, reproduced or transmitted without prior
permission of the author



Researcher declaration to accompany the submission of written work

I Dominik Thaler certify that I am the author of the work *Essays on Macroeconomics with Financial Frictions* I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297)).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Statement of inclusion of previous work (if applicable):

I confirm that chapter 1 was jointly co-authored with Angela Abbate and I contributed 50% of the work.

Signature and Date:

Dominik Thaler

29/07/2016

Abstract

The first chapter of this thesis, joint with Angela Abbate analyses the importance of the risk-taking channel for monetary policy. To answer this question, we develop and estimate a quantitative monetary DSGE model where banks choose excessively risky investments, due to an agency problem which distorts banks' incentives. As the real interest rate declines, these distortions become more important and excessive risk taking increases, lowering the efficiency of investment. We show that this novel transmission channel generates a new and quantitatively significant monetary policy trade-off between inflation and real interest rate stabilization: it is optimal for the central bank to tolerate greater inflation volatility in exchange for lower risk taking.

The second chapter develops a quantitative model of sovereign default with endogenous default costs to propose a novel answer to the question why governments repay their debt. In the model domestic banks are exposed to sovereign debt. Hence sovereign default causes large losses for the banks, which translate into a financial crisis. The government trades these costs off against the advantage of not repaying international investors. Besides replicating business cycle moments, the model is able to generate not only output costs of a realistic magnitude, but also endogenously predicts that default is followed by a period during which no new foreign lending takes place. The duration of this period matches empirical estimates.

The third chapter outlines a method to reduce the computationally necessary state space for solving dynamic models with global methods. The idea is to replace several state variables by a summary state variable. This is made possible by anticipating future choices that depend on one of the replaced variables. I explain how this method can be applied to a simple portfolio choice problem.

für meinen Opi Gerhard Burret

Acknowledgements

I am very grateful to my supervisors Evi Pappa and Árpád Ábrahám, who have guided and supported me and my research patiently and constantly throughout my PhD. Special thanks go to Enrique Mendoza, Leopold von Thadden and Raf Wouters for hosting me at their institutions and discussing my research with me. I am very grateful to Angela Abbate, whose optimistic attitude made work our common project a lot easier. I would also like to thank Luisa Lambertini and Peter Karadi for being part of the defense committee.

During the five years I spent at the EUI I made many friends at the department whose contribution to this thesis not only consists in the many things they helped me understand, but also in celebrating the good life Florence offers outside the walls of the institute: Omar, Andreu, Rana, Guilherme, Clement, Vincent, Brais, Paveł, Luis, Zelda, John. The latter is equally true for my flatmates Ledda and Emi, who became my loving family for my first year in Florence.

Most of all I thank Mireia and my parents for the infinite support that carried me through to the PhD.

Contents

I	Monetary policy an the asset risk taking channel	1
1	Introduction	1
2	A Dynamic New Keynesian model with a bank risk-taking channel	4
2.1	Households	4
2.2	Equity and deposit funds	5
2.3	Capital producers	6
2.4	The Bank	7
2.4.1	Second-stage problem	9
2.4.2	First-stage problem	10
2.4.3	Closing the bank model: the zero-profit condition	10
2.4.4	Properties of the banking sector equilibrium	11
2.4.5	Full model with deposit insurance and liquidation value	14
2.5	Labour and goods sectors	16
2.6	Monetary and fiscal policy	16
3	Dynamic implications of the risk-taking channel in the estimated model	16
3.1	Model estimation	16
3.2	Dynamic implications of excessive risk taking	17
3.3	Evaluating the fit of the estimated model	19
4	Monetary policy with the risk-taking channel	21
4.1	The central bank problem	22
4.2	Findings	23
5	Conclusion	25
	References	26
	Appendices	29
II	Austerity to save the banks? A quantitative model of sovereign default with endogenous default costs and a financial sector	49
1	Introduction	49

2	The model	52
2.1	The household	52
2.2	The banks	53
2.3	The firms	57
2.4	The foreign investors	58
2.5	Market clearing	58
2.6	The government	59
2.7	Discussion of key assumptions	62
3	Financial frictions, sovereign debt, and the costs of default	64
3.1	Underinvestment due to the financial friction	64
3.2	A new motive for international borrowing	65
3.3	The costs of default	66
3.4	Default incentives	69
4	Computation and calibration	73
4.1	Computation	73
4.2	Calibration	75
5	Quantitative results	78
5.1	Cyclical moments	78
5.2	Default episodes	80
5.3	Robustness	83
6	Policy application: The EU plan to reduce the exposure of banks to domestic sovereign debt	84
7	Conclusion	85
	References	87
	Appendices	91
III	Reducing the computationally necessary state space by anticipating future choices	95
1	Introduction	95
2	Example: A neoclassical growth model	97
3	Applicability of the method	102
	References	104

Chapter I

Monetary policy and the asset risk taking channel

joint with Angela Abbate

1 Introduction

The recent financial crisis has sparked a debate about the influence of monetary policy on the risk-taking behaviour of the banking sector. A number of recent studies such as Jimenez et al. (2014) show that low interest rates increase the risk appetite of banks, creating an additional channel of monetary policy transmission, known as the risk-taking channel.¹ Though there has been much discussion of the risk-taking channel amongst policy makers in recent years,² its general-equilibrium and optimal monetary-policy implications remain unclear. Answering these questions requires a quantitative model which is consistent with both the evidence on the risk-taking channel and with conventional views about monetary policy. Our contribution is to build and estimate a medium-scale New Keynesian DSGE model, where monetary policy influences bank risk taking, which in turn affects the real economy. Furthermore, we provide analytical results which show how the inefficiency of risk taking depends on the volatility of the real interest rate, implying a motive for the policy maker to stabilize the real interest rate, at the cost of greater inflation volatility. This constitutes a new trade-off that influences optimal monetary policy in a quantitatively significant way.

In our model banks raise funds through deposit and equity, which they then use to invest in risky capital projects. In particular, banks choose from a continuum of investment projects, each defined by different risk-return characteristics. Every project has a certain probability of being successful and yielding capital in the next period. However, the safer the project, the lower the return in the event of success. As in Dell'Ariccia et al. (2014), we assume that depositors cannot observe the investment risk choice and that bank owners are protected by limited liability. These two assumptions create an agency problem:

The authors would like to thank Arpad Abraham, Pelin Ilbas, Ester Faia, Stephane Moyon, Evi Pappa, Raf Wouters, our colleagues at the Bundesbank, European University Institute and National Bank of Belgium, as well as participants to the Applied Time Series Econometrics Workshop in St. Louis, XX Workshop on Dynamic Macroeconomics in Vigo and to the 6th Bundesbank-CFS-ECB Workshop on Macro and Finance in Frankfurt for useful comments. We thank the NBB for funding Dominik Thaler's research stay in Brussels.

¹This term was first used by Borio and Zhu (2008).

²See, for instance, New York Times, 'Fed officially concedes risk of low rates but signals no shift', 11.10.2010, Financial Times, 'Draghi warns central banks against blind risk taking', 14.5.2015, Bloomberg Business, 'Carney says QE can encourage excessive risk taking in financial markets', 25.1.2015.

banks are partially isolated from the downside risk of their investment and choose a risk level that is socially excessive.³ The agency problem could be mitigated if bankers held more equity. Yet, banks rely on both types of funding because equity is relatively more costly than deposits, as a result of deposit insurance and a friction in the equity market. A lower risk-free rate increases the relative cost advantage of deposits. As a consequence, banks respond by leveraging up and choosing riskier investment projects. This higher risk implies a lower average efficiency of investment, which leads to a decline in the capital stock. Our model hence generates a new transmission channel through which monetary policy affects the real economy. This channel dampens the positive effects of expansionary monetary policy, as reductions in the interest rate exacerbate the financial market distortions and their implied inefficiency. We validate the quantitative implications of the model by estimating it on US data. Posterior odds show that the inclusion of the risk-taking channel improves the in-sample fit for nonfinancial variables. At the same time our model predicts a path of risk taking that matches survey evidence on the riskiness of newly issued loans from the Fed Survey of Terms of Business Lending.

We use the model to analyse the normative implications of the risk-taking channel, and their quantitative importance in terms of consumer welfare. While in standard New Keynesian models it is optimal for the central bank to stabilize inflation,⁴ the risk-taking channel introduces a motive for real interest rate stabilization. We show theoretically that the efficiency of the banks' risk choice depends negatively on the volatility of the real interest rate, thereby providing a motive for stabilizing the real interest rate. The risk-taking channel therefore generates a new monetary policy trade-off, as this additional objective conflicts with inflation stabilization. Moreover, we show that this new objective alters optimal policy in a quantitatively significant way. Using optimal simple rules, we find that the central bank optimally accepts around 50% more inflation volatility relative to the case without the channel, in return for a more stable real interest rate. Furthermore, ignoring the risk-taking channel comes at a welfare cost equivalent to a 0.5-1.0% loss in lifetime consumption. These results contradict existing findings for other general-equilibrium models with financial frictions such as De Fiore and Tristani (2013) and Bernanke and Gertler (2001), where optimal policy remains close to full price stability even if financial frictions are introduced. In particular, De Fiore and Tristani (2013) characterize Ramsey policy in a small New Keynesian set-up with credit frictions, where firms borrow in advance to pay wages, and where default risk and costly monitoring generate a spread between the loan rate and the risk-free rate. The authors show that the presence of credit frictions augments the otherwise standard second-order approximation of the welfare function with one additional term: i.e. that interest-rate and credit-spread volatility

³By socially excessive we mean that it exceeds the risk level that would be chosen, if no friction were present.

⁴Though in standard New Keynesian models a trade-off between output and inflation stabilization is generated by cost push shocks, the relative importance of output stabilization is typically found to be small.

directly influence welfare. However, this additional term is found to be quantitatively small, so that optimal policy does not substantially deviate from price stability.

Our work relates to a growing theoretical literature that links monetary policy to financial sector risk in a general-equilibrium framework. Yet, several features distinguish our work from existing ones. First, motivated by the evidence reviewed above, it is the first to explicitly model the effect of monetary policy on the riskiness of banks' assets and its macroeconomic effects. Second, we show theoretically how the risk-taking channel generates a new significant trade-off for the monetary policy authority. Most of the existing literature explores risk on the funding side of banks' balance sheets, associating risk with increased leverage. For instance, several models build on the financial accelerator framework of Bernanke et al. (1999).⁵ The mechanism in these models relies on the buffer role of equity, and therefore leverage is found to be counter-cyclical with respect to the balance sheet size. Our model, by contrast, gives rise to pro-cyclical leverage, which is in line with the empirical evidence reported in Adrian and Shin (2014) and Adrian et al. (2015). Another example is Angeloni and Faia (2013) and Angeloni et al. (2015), where lower interest rates translate into a higher bank leverage, and a higher fraction of inefficient bank runs. Asset risk, on the other hand, has so far mainly been discussed in the literature on optimal regulation such as Christensen et al. (2011) and Collard et al. (2012). In these papers, however, either the depositors or the financial regulator ensure that risk is always chosen optimally, so monetary policy has no influence on risk taking.⁶ In contrast to the previous two papers, we provide micro-foundations for the asset risk-taking channel and focus on monetary policy while abstracting from regulation.⁷

Our model of the asset risk-taking channel explains two stylized facts documented by recent empirical evidence. First, low interest rates cause banks to make riskier investments. Using micro data from the Spanish Credit Register, Jimenez et al. (2014) find that lower interest rates induce banks to make relatively more loans to firms that qualify as risky *ex ante* as well as *ex post*.⁸ Second, the increase in risk taking is not fully compensated for by higher risk premia, as shown by Buch et al. (2014) and Ioannidou et al. (2014). As a consequence, the expected return on banks' investment decreases, as risk increases in response to lower interest rates. Moreover, the model posits that the bank risk choice is determined by the level of leverage, rather than the quantity of loans: a modelling choice which is in line with the findings of Ioannidou et al. (2014) and Jimenez

⁵For example, in Gertler et al. (2012) and de Groot (2014) a monetary expansion increases banking sector leverage, which in turn amplifies the financial accelerator and strengthens the propagation of shocks to the real economy.

⁶Both papers feature ad-hoc extensions that relate risk to the amount of lending and hence indirectly to monetary policy.

⁷One could reinterpret our model as applying to an economy where regulation is unable to fully control risk taking.

⁸This finding is confirmed by Dell'Ariccia et al. (2013), Angeloni et al. (2015), Afanasyeva and Guentner (2015) and Buch et al. (2014) for the US and by Ioannidou et al. (2014) for Bolivia. We complement these results in appendix D, where we show that further times series evidence on the risk-taking channel for the US banking sector based on data from the terms of business lending survey.

et al. (2014).

The paper is structured as follows. In Section 2 we develop a DSGE model of the asset risk-taking channel. Section 3 presents the results from the estimation of the model and discusses the dynamic implications of bank risk taking. Section 4 analyzes how monetary policy should be conducted if the risk-taking channel is present, and Section 5 concludes.

2 A Dynamic New Keynesian model with a bank risk-taking channel

We build a general-equilibrium model where competitive banks obtain funds from depositors and equity holders, and invest them into capital projects executed by capital producers. Every bank chooses its investment from a continuum of technologies, each defined by a given risk-return characteristic. The risk choice of the bank is distorted by an agency problem and affected by the level of the real interest rate. This model reproduces two features found in the data: risk taking depends on the contemporaneous interest rate and is priced inefficiently. The real sector of the economy features otherwise standard elements, which are therefore sketched only briefly here. More details on the standard sectors and the complete set of equations characterizing the model can be found in Appendix B.

2.1 Households

The representative household chooses consumption c_t , working hours L_t and savings in order to maximize its discounted lifetime utility. Saving is possible through three instruments: government bonds s_t , which pay the safe gross nominal interest rate R_t , deposit funds d_t , and bank equity funds e_t . The two funds enable the household to invest into the banking sector, and pay an uncertain nominal return of $R_{d,t+1}$ and $R_{e,t+1}$.⁹ Maximization of his lifetime utility (see Appendix B) yields the usual labour supply condition, the Euler equation, and two no-arbitrage conditions:

$$E_t \left[\Lambda_{t+1} \frac{R_{d,t+1}}{\pi_{t+1}} \right] = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] , \quad (1)$$

$$E_t \left[\Lambda_{t+1} \frac{R_{e,t+1}}{\pi_{t+1}} \right] = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] , \quad (2)$$

where Λ_t is the marginal utility of consumption.

⁹In our notation the time index refers to the period when a variable is determined.

2.2 Equity and deposit funds

As we explain in detail below, there is a continuum of banks which intermediate the households' savings using deposits and equity. Each bank is subject to a binary idiosyncratic shock which makes a bank fail with probability $1 - q_{t-1}$, in which case equity is wiped out completely and depositors receive partial compensation from the deposit insurance scheme. We assume that households invest into bank equity and deposits through two funds, which diversify away the idiosyncratic bank default risk by investing into all banks.¹⁰

The deposit fund works without frictions, and represents the depositors' interests perfectly. It raises money from the households and invests it into d_t units of deposits.¹¹ In the next period, the fund receives the nominal deposit rate $r_{d,t}$ from each bank that does not fail. Deposits of failing banks are partially covered by deposit insurance. Most deposit insurance schemes around the world, including the US, guarantee all deposits up to a certain maximum amount per depositor.¹² We model this capped insurance scheme by assuming that the deposit insurance guarantees deposits up to a fraction ψ of total bank liabilities $e_t + d_t$. We assume that the deposit insurance cap is inflation-adjusted, to avoid complicating the monetary policy trade-off by allowing an interdependence between monetary policy and deposit insurance. As we will show later, the deposit insurance cap is always binding in equilibrium, i.e. the bank's liabilities exceed the cap of the insurance $r_{d,t}d_t > \psi(d_t + e_t)\pi_{t+1}$. Defining the equity ratio $k_t = \frac{e_t}{d_t + e_t}$, the deposit fund therefore receives a real return of $\psi/(1 - k_t)$ per unit of deposits from each defaulting bank at t . The deposit fund hence pays a nominal return of:

$$R_{d,t+1} \equiv q_t r_{d,t} + (1 - q_t) \frac{\psi}{1 - k_t} \pi_{t+1} . \quad (3)$$

Unlike the deposit fund, the equity fund is subject to a simple agency problem. In particular, we assume that the fund manager faces two options. He can behave diligently and use the funds raised at t to invest into e_t units of bank equity. A fraction q_t of banks pay back a return of $r_{e,t+1}$ next period, while defaulting banks pay nothing. Alternatively, the fund manager can abscond with the funds and consume a fraction ξ_t in the subsequent period, while the rest is lost. To prevent the fund manager from doing so, the equity providers promise to pay him a premium p_t at time $t + 1$ conditional on not absconding. Equity providers pay the minimal premium that induces diligent behaviour, i.e. $p_t = \xi_t e_t$. This premium is rebated to the household in a lump-sum fashion. Once absconding is ruled out in equilibrium, the equity fund manager perfectly represents the interests of

¹⁰Focusing on idiosyncratic risk is a simplification that keeps the model tractable. Since more idiosyncratic risk is undesirable (as we will see below), assuming instead that bank risk also has an aggregate dimension would just reinforce the results.

¹¹We use deposits to refer to both units of deposit funds and units of bank deposits since they are equal. We do the same for equity.

¹²For a comprehensive documentation see, for instance, Demirgüç-Kunt et al. (2005).

the equity providers. The equity fund hence pays the return on bank equity net of the premium:

$$R_{e,t+1} \equiv q_t r_{e,t+1} - \xi_t \pi_{t+1} . \quad (4)$$

We allow the equity premium ξ_t to vary over time.¹³ Since bank equity is the residual income claimant, the return on the equity fund is affected by all types of aggregate risk that influences the return of surviving banks.

The two financial distortions introduced so far have important implications. The agency problem implies an equity premium, i.e. a premium of the risk-adjusted return on equity over the risk-free rate. Deposit insurance, on the other hand, acts as a subsidy on deposits, which implies a discount on the risk-adjusted return on deposits. As explained below, the difference in the costs of these two funding types induces a meaningful trade-off between bank equity and bank deposits under limited liability.

2.3 Capital producers

We assume that the capital production process is risky in a way that nests the standard capital production process in the New Keynesian model. In particular, capital is produced by a continuum of capital producers indexed by m . At period t they invest i_t^m units of final good into a capital project of size o_t^m . This project is successful with probability q_t^m , in which case the project yields $(\omega_1 - \frac{\omega_2}{2} q_t^m) o_t^m$ units of capital at $t + 1$. Otherwise, the project fails and only the liquidation value of θo_t^m units of capital can be recovered (where $\theta \ll \omega_1 - \frac{\omega_2}{2} q_t^m$). Each capital producer has access to a continuum of technologies with different risk-return characteristics indexed by $q^m \in [0, 1]$. Given a chosen technology q_t^m , the output of producer m is therefore:

$$K_t^m = \begin{cases} \left(\omega_1 - \frac{\omega_2}{2} q_t^m \right) o_t^m & \text{with probability } q_t^m \\ \theta o_t^m & \text{else} \end{cases}$$

This implies that the safer the technology (higher q_t^m), the lower output in the event of success.

The bank orders capital projects, and requires the capital producer to use a certain technology, but this choice cannot be observed by any third party. Given the technology choice q_t , and assuming that the projects of individual producers are uncorrelated, we can exploit the law of large numbers to derive aggregate capital:

$$K_t = o_t \left(q_t \left(\omega_1 - \frac{\omega_2}{2} q_t \right) + (1 - q_t) \theta \right) . \quad (5)$$

Furthermore we assume that capital, which depreciates at rate δ , becomes a project

¹³This shock, driving a wedge between deposit and safe rates on one hand, and equity rates on the other, is similar to the risk premium shock often found in medium-scale DSGE models (e.g. Smets and Wouters (2007)). Like all shocks in the model, it follows a standard lognormal AR(1) process.

(of undefined q_t) at the end of every period. That is, existing capital may be destroyed due to unsuccessful reuse, and it can be reused under a different technology than it was originally produced.¹⁴

The total supply of capital projects by the capital producers is the sum of the existing capital projects $o_t^{old} = (1 - \delta)K_{t-1}$, which they purchase from the owners (the banks) at price Q_t , and the newly created projects o_t^{new} , which are created by investing i_t units of the final good. We allow for investment adjustment costs and investment efficiency shocks, i.e. we assume that i_t units of investment yield $\varepsilon_t^I (1 - S(i_t/i_{t-1}))$ units of project, where $S = \kappa \left(\frac{i_t}{i_{t-1}} - 1 \right)^2$. Hence $o_t = o_t^{new} + o_t^{old}$ and $o_t^{new} = \varepsilon_t^I \left(1 - S \left(\frac{i_t}{i_{t-1}} \right) \right) i_t$. Capital producers maximize their expected discounted profits taking as given the price Q_t and the household's stochastic discount factor:¹⁵

$$\max_{i_t, o_t^{old}} E_t \sum_0^\infty \beta^t \Lambda_t \left[Q_t \varepsilon_t^I \left(1 - S \left(\frac{i_t}{i_{t-1}} \right) \right) i_t + Q_t o_t^{old} - i_t - Q_t o_t^{old} \right].$$

While the old capital projects are always reused, the marginal capital project is always a new one.¹⁶ Hence, the price of projects Q_t is determined by new projects according to the well known Tobin's q equation:

$$Q_t \varepsilon_t^I \left[1 - S \left(\frac{i_t}{i_{t-1}} \right) - S' \left(\frac{i_t}{i_{t-1}} \right) \frac{i_t}{i_{t-1}} \right] - 1 = \beta E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \varepsilon_{t+1}^I Q_{t+1} S' \left(\frac{i_{t+1}}{i_t} \right) \left(\frac{i_{t+1}}{i_t} \right)^2 \right]. \quad (6)$$

Note that our model of risky capital production boils down to the standard riskless setting of the New Keynesian model if we fix $q_t = \bar{q}$ and choose parameters such that $q_t \left(\omega_1 - \frac{\omega_2}{2} q_t \right) + (1 - q_t)\theta = 1$.

2.4 The Bank

The bank is the central agent of our model: it raises resources through deposits and equity and invests them into a risky project. As in Dell'Ariccia et al. (2014), an agency problem arises between banks and depositors when choosing the risk level, since depositors cannot observe the banks' risk choice and banks are protected by limited liability. The less equity a bank has, the higher the incentives for risk taking. Yet, since deposit insurance and the equity premium drive a wedge between the costs of deposits and equity, the banks' optimal capital structure comprises both equity and deposits, balancing the agency problem associated with deposits with the higher costs of equity. We will show that the equilibrium risk chosen by the banks is excessive, and that the interest rate influences the

¹⁴This assumption ensures that we do not have to keep track of the distribution of different project types. Think of a project as a machine yielding capital services, which can be run at different speeds (levels of risk). In case it is run at a higher speed, the probability of an accident destroying the machine is higher. After each period the existing machines are overhauled by the capital producers and at this point the speed setting can be changed.

¹⁵Their out-of-steady-state profits are rebated lump sum to the household.

¹⁶We abstract from a non-negativity constraint on new projects.

degree of its excessiveness.

We assume that there is a continuum of banks which behave competitively so that there is a representative bank (we therefore omit the bank's index in what follows). The bank is owned by the equity providers, and hence maximizes the expected discounted value of profits¹⁷ using the household's stochastic discount factor. Every period, the bank optimally chooses its liability structure by raising deposits d_t and equity e_t from the respective funds. These resources are then invested into o_t capital projects, purchased at price Q_t . When investing into capital projects, the bank chooses the risk characteristic q_t of the technology applied by the capital producer. This risk choice is not observable for depositors. Each bank can only invest into one project and hence faces investment risk:¹⁸ with probability q_t the bank receives a high payoff from the capital project; with probability $1 - q_t$ the investment fails and yields only the liquidation value. Assuming a sufficiently low liquidation value θ , a failed project implies the default of the bank. In this case, given limited liability, equity providers get nothing and depositors get the deposit insurance benefit. In case of success the bank can repay its investors: depositors receive their promised return $r_{d,t}$ and equity providers get the state-contingent return $r_{e,t+1}$.

It is useful to think of the bank's problem as a recursive two-stage problem. At the second stage, the bank chooses the optimal risk level q_t given a certain capital structure and a certain cost of deposits. At the first stage, the bank chooses the optimal capital structure, anticipating the implied solution for the second-stage problem. Note that not only the bank but also the bank's financiers anticipate the second-stage risk choice and price deposits and equity accordingly, which is understood by the bank.

Before we derive the solution for this recursive problem, we establish the bank's objective function. Per dollar of nominal funds raised in period t the bank purchases $1/(Q_t P_t)$ units of the capital project from the capital producer, choosing a certain riskiness q_t . If the project is successful it turns into $(\omega_1 - \frac{\omega_2}{2} q_t)/(Q_t P_t)$ capital goods. In the next period $t + 1$, the bank rents the capital to the firm, which pays the real rental rate $r_{k,t+1}$ per unit of capital. Furthermore the bank receives the depreciated capital, which becomes a capital project again, with a real value of $(1 - \delta)Q_{t+1}$ per unit of capital. The bank's total nominal income, per dollar raised, conditional on success is therefore:

$$\left(\omega_1 - \frac{\omega_2}{2} q_t\right) \frac{r_{k,t+1} + (1 - \delta)Q_{t+1}}{Q_t} \frac{P_{t+1}}{P_t}.$$

¹⁷Profits in excess of the opportunity costs of equity.

¹⁸The assumption that the bank can only invest into one project and cannot diversify the project risk might sound stark. Yet three clarifications are in place: First, our set up is isomorphic to a model where the bank invests into an *optimally* diversified portfolio of investments but is too small to *perfectly* diversify its portfolio. The binary payoff is then to be interpreted as the portfolio's expected payoff conditional on default or repayment respectively. Second, if the bank was able to perfectly diversify risk, then limited liability would become meaningless and we would have a model without financial frictions. Third, we don't allow the bank to buy the safe asset. Yet this assumption is innocuous: since the banks demand a higher return on investment than the households due to the equity premium, banks wouldn't purchase the safe asset even if they could.

At the same time, the bank has to repay the deposit and equity providers. Using the equity ratio k_t , the total nominal repayment per dollar of funds due in $t + 1$ in the event of success is $r_{e,t+1}k_t + r_{d,t}(1 - k_t)$.

The bank maximizes the expected discounted value of excess profits, i.e. revenues minus funding costs, using the stochastic discount factor of the equity holders, i.e. the household. Given the success probability of q_t ,¹⁹ the bank's objective function is:

$$\max_{q_t, k_t} \beta E \left[\frac{\Lambda_{t+1}}{\pi_{t+1}} q_t \left(\left(\omega_1 - \frac{\omega_2}{2} q_t \right) \frac{r_{k,t+1} + (1 - \delta) Q_{t+1}}{Q_t} \pi_{t+1} - r_{d,t}(1 - k_t) - r_{e,t+1}k_t \right) \right]. \quad (7)$$

Note that we did not multiply the per-unit profits by the quantity of investment. By doing so we anticipate the equilibrium condition that the bank, whose objective function is linear in the quantity of investment, needs to be indifferent about the quantity of investment. The quantity will be pinned down together with the return on capital by the bank's balance sheet equation $e_t + d_t = Q_t o_t$, the market clearing and zero-profit conditions.

The bank's problem can be solved analytically, yet the expressions get fairly complex. Therefore we derive here the solution for $\psi = \theta = 0$, that is without deposit insurance and with a liquidation value of 0. This simplifies the expressions but the intuition remains the same. Allowing ψ and θ to be nonzero on the other hand is necessary to bring the model closer to the data. The solution for the general case is discussed in section 3.5.5.

To make notation more tractable we rewrite the bank's objective function (7) in real variables expressed in marginal utility units:²⁰

$$\omega_1 q_t \tilde{r}_{l,t} - \frac{\omega_2}{2} q_t^2 \tilde{r}_{l,t} - q_t \tilde{r}_{d,t}(1 - k_t) - q_t \tilde{r}_{e,t} k_t, \quad (8)$$

For later use we rewrite the household's no-arbitrage conditions (1) and (2) combined with the definition of the funds' returns (3) and (4) as $\tilde{r}_{d,t} = \frac{\tilde{R}_t}{q_t}$ and $\tilde{r}_{e,t} = \frac{\tilde{R}_t + \tilde{\xi}_t}{q_t}$. We now solve the bank's problem recursively.

2.4.1 Second-stage problem

At the second stage, the bank has already raised $e_t + d_t$ funds and now needs to choose the risk characteristic of the investment q_t , such that equity holders' utility is maximized. As already mentioned, we assume that the bank cannot write contracts conditional on q_t with the depositors at stage one, since q_t is not observable to them. Therefore, at the second stage the bank takes the deposit rate as given. Furthermore, since the capital

¹⁹Here we anticipate that the bank defaults in the event of a bad project outcome. See section 2.4.5

²⁰That is, we use the following definitions: $\tilde{r}_{l,t} = E_t \left[\Lambda_{t+1} \left(\frac{r_{k,t+1} + (1 - \delta) Q_{t+1}}{Q_t} \right) \right]$, $\tilde{r}_{d,t} = E_t \left[\Lambda_{t+1} \frac{r_{d,t}}{\pi_{t+1}} \right]$, $\tilde{r}_{e,t} = E_t \left[\Lambda_{t+1} \frac{r_{e,t+1}}{\pi_{t+1}} \right]$, $\tilde{R}_t = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right]$, $\tilde{\xi}_t = E_t [\Lambda_{t+1} \xi_t]$.

structure is already determined, maximizing the excess profit coincides with maximizing the profit of equity holders. The bank's objective function is therefore:

$$\max_{q_t} \omega_1 q_t \tilde{r}_{l,t} - \frac{\omega_2}{2} q_t^2 \tilde{r}_{l,t} - q_t \tilde{r}_{d,t} (1 - k_t) . \quad (9)$$

Deriving problem (9) with respect to q_t yields the following first-order condition:²¹

$$q_t = \frac{\omega_1 \tilde{r}_{l,t} - \tilde{r}_{d,t} (1 - k_t)}{\omega_2 \tilde{r}_{l,t}} . \quad (10)$$

2.4.2 First-stage problem

At the point of writing the deposit contract at stage one, depositors anticipate the bank's choices at stage two and therefore the depositors' no-arbitrage condition $\tilde{r}_{d,t} = \frac{\tilde{R}_t}{q_t}$ must hold in equilibrium.²² Using this equation together with the previous first-order condition (10) we can derive the optimal q_t as a function of k_t and $\tilde{r}_{l,t}$:

$$\hat{q}_t \equiv q_t(k_t) = \frac{1}{2\omega_2 \tilde{r}_{l,t}} \left(\omega_1 \tilde{r}_{l,t} + \sqrt{(\omega_1 \tilde{r}_{l,t})^2 - 4\omega_2 \tilde{r}_{l,t} \tilde{R}_t (1 - k_t)} \right) . \quad (11)$$

We can now solve the first-stage problem of the bank, which chooses the capital structure k_t to maximize excess profits, anticipating the $q_t(k_t)$ that will be chosen at the second stage:

$$\max_{k_t} \hat{q}_t \omega_1 \tilde{r}_{l,t} - \frac{\omega_2}{2} \tilde{r}_{l,t} \hat{q}_t^2 - q_t \tilde{r}_{d,t} (1 - k_t) - q_t \tilde{r}_{e,t} k_t , \quad (12)$$

subject to the no-arbitrage condition for depositors ($\tilde{r}_{d,t} = \frac{\tilde{R}_t}{q_t}$) and for equity providers ($\tilde{r}_{e,t} = \frac{\tilde{R}_t + \tilde{\xi}_t}{q_t}$). Plugging these in and deriving, we obtain the first-order condition for k_t :

$$\omega_1 \tilde{r}_{l,t} \frac{\partial \hat{q}_t}{\partial k_t} - \tilde{\xi}_t - \frac{\omega_2}{2} \tilde{r}_{l,t} \frac{\partial \hat{q}_t^2}{\partial k_t} = 0 . \quad (13)$$

which (assuming an interior solution) can be solved for k_t as:

$$\hat{k}_t \equiv k_t(\tilde{r}_{l,t}) = 1 - \frac{\tilde{\xi}_t (\tilde{R}_t + \tilde{\xi}_t) (\omega_1 \tilde{r}_{l,t})^2}{\omega_2 \tilde{R}_t \tilde{r}_{l,t} (\tilde{R}_t + 2\tilde{\xi}_t)^2} . \quad (14)$$

2.4.3 Closing the bank model: the zero-profit condition

Since there is a continuum of identical banks, each bank behaves competitively and takes the return on investment $\tilde{r}_{l,t}$ as given. Perfect competition and free entry imply that banks will enter until there are no expected excess profits to be made. In the presence

²¹We focus on interior solutions and choose the larger of the two roots, which is the closest to the optimum, as we will see below.

²²Note that the agency problem arises from the fact that the bank does not consider this as a constraint of its maximization problem.

of uncertainty it is natural to focus on the case that banks make no excess profit in any future state of the world:

$$\left(\omega_1 - \frac{\omega_2}{2}q_{t-1}\right) \left(\frac{r_{k,t} + (1-\delta)Q_t}{Q_{t-1}}\right) - \frac{r_{d,t-1}}{\pi_t}(1 - \hat{k}_{t-1}) - \frac{r_{e,t}}{\pi_t}\hat{k}_{t-1} = 0 . \quad (15)$$

Using the equity and deposit supply schedules and taking expectation over this equation we get:

$$\hat{q}_t\omega_1\tilde{r}_{l,t} - \frac{\omega_2}{2}\tilde{r}_{l,t}\hat{q}_t^2 - \hat{k}_t\tilde{\xi}_t - R_t = 0 . \quad (16)$$

Combining (16) with the optimality conditions (11) and (14), we can derive analytical expressions for the equity ratio k_t , riskiness choice q_t (the last term in each row is an approximation under certainty equivalence and $R_t^r \equiv R_t/E_t[\pi_{t+1}]$):

$$k_t = \frac{\tilde{R}_t}{\tilde{R}_t + 2\tilde{\xi}_t} \approx \frac{R_t^r}{R_t^r + 2\xi_t} \quad (17)$$

$$q_t = \frac{\omega_1(\tilde{\xi}_t + \tilde{R}_t)}{\omega_2(2\tilde{\xi}_t + \tilde{R}_t)} \approx \frac{\omega_1(\xi_t + R_t^r)}{\omega_2(2\xi_t + R_t^r)} \quad (18)$$

2.4.4 Properties of the banking sector equilibrium

These results for the banking sector risk choice have five interesting implications that we first summarize in a proposition, before intuitively discussing them in turn.

Proposition 1: Be $[\tilde{r}_{l,t}, q_t, k_t]$ an equilibrium in the banking sector with interior bank choices under perfect competition. Denote the expected return on investment in capital units by $f(q_t) \equiv \left(\omega_1 - \frac{\omega_2}{2}q_t\right)q_t$. Then:

- (1) Risk decreases in the real interest rate: $\frac{\partial q_t}{\partial \tilde{R}_t} > 0$.
- (2) The equity ratio increases in the real interest rate: $\frac{\partial k_t}{\partial \tilde{R}_t} > 0$.
- (3) Risk taking is excessive: $q_t < \arg\max f(q_t)$.
- (4) The expected return on investment increases in the real interest rate: $\frac{\partial f(q_t)}{\partial \tilde{R}_t} > 0$.
- (5) The expected return on investment is a concave function of the real interest rate $\frac{\partial^2 f(q_t)}{\partial \tilde{R}_t^2} < 0$.

The proof can be found in Appendix C.

The first two results can be easily seen from equations (17) and (18). As the real risk-free rate R_t^r decreases, the equity ratio k_t falls as banks substitute equity with deposits and

the riskiness of the bank increases (q_t falls).²³ The intuition behind this result is as follows: On the one hand, a lower risk-free rate decreases the rate of return on capital projects, reducing the benefits of safer investments, conditional on repayment. This induces the bank to adopt a riskier investment technology. On the other hand, the lower risk-free rate reduces the cost of funding, leaving more resources available to the bank's owners in case of repayment: this force contrasts with the first one, making safer investments more attractive. There is a third force: a lower risk-free interest rate means that the equity premium becomes relatively more important. As a result the bank shifts from equity to deposits, internalizing less the consequences of the risk decision and choosing a higher level of risk. The first and third effects dominate, and overall a decline in the real interest rate induces banks to choose more risk. Note that these two results depend on the assumption that the (discounted) equity premium is independent of the (discounted) real interest rate. If we allowed the equity premium to be a function $\tilde{\xi}_t(\tilde{R}_t)$ of the real interest rate, the result would continue to hold under the condition that $\tilde{\xi}_t(\tilde{R}_t) > \tilde{\xi}_t'(\tilde{R}_t) \tilde{R}_t$, which rules out proportionality. This mechanism provides a rationalization of the empirical finding in section 2: that a decline in the nominal interest rate²⁴ causes an increase in bank risk-taking behaviour.

The third result implies that the bank's investment could have a higher expected return (in units of capital) if the bank chose a higher level of safety. In other words, risk taking is excessive, i.e. suboptimally high. This is due to the agency problem, which arises from limited liability and the lack of commitment/contractability of the banker regarding his risk choice. The importance of this friction can be assessed by comparing the solution of the imperfect markets bank model with the solution of the model without any frictions. The frictionless risk choice can be derived under any of the following alternative scenarios: either both the equity premium and deposit insurance are zero (which eliminates the cost disadvantage of equity and leads to 100% equity finance), or contracts are complete and deposit insurance is zero (which eliminates the agency problem and leads to 100% deposit finance), or liability is not limited and deposit insurance is zero (as before), or households invests directly into a diversified portfolio of capital projects (which eliminates the financial sector altogether). Since in a frictionless model q_t is chosen to maximize the consumption value of the expected return:

$$\max_{q_t} \tilde{r}_{l,t} \left(\omega_1 - \frac{\omega_2}{2} q_t \right) q_t ,$$

the optimal level of q_t trivially is $q_t^o = \frac{\omega_1}{\omega_2}$. Comparing the frictionless risk choice q_t^o and the choice given the friction q_t^f

²³At least under certainty equivalence or up to a first-order approximation, when the Λ_{t+1} terms contained in the tilde variables cancel each other out.

²⁴In a monetary model, a cut in the nominal interest rate, the standard monetary policy tool, is followed by a decline in the real interest rate due to price stickiness.

$$q_t^f = q_t^o \frac{\tilde{\xi}_t + \tilde{R}_t}{2\tilde{\xi}_t + \tilde{R}_t} \approx q_t^o \frac{\xi_t + R_t^r}{2\xi_t + R_t^r},$$

we observe that the agency friction drives a wedge between the frictionless risk level and the level that is actually chosen. This wedge has two important features. First, it is smaller than one,²⁵ implying that under the agency problem the probability of repayment is too low, and hence banks choose excessive risk. Second, note that the wedge depends on R_t^r and that the derivative of the first-order approximation of the wedge w.r.t. R_t^r is positive. This implies that the wedge increases, i.e. risk taking gets more excessive, as the real interest rate falls. As we move further away from the optimal level of risk the expected return on investments necessarily falls, which is the fourth result above.

But it is not only the bank risk choice that is suboptimal. The capital structure is chosen suboptimally too. If banks could commit to choose the optimal level of risk, they would not need any skin in the game. Hence they would avoid costly equity and would finance themselves entirely through deposits: $k_t^o = 0$. Instead they choose $k_t^f = \frac{\tilde{R}_t}{\tilde{R}_t + 2\tilde{\xi}_t}$. The equity ratio resembles the two features of the risk taking. First, there is excessive use of equity funding. Second, the equity ratio is increasing in R_t^r up to a first-order approximation.

Both the risk and the capital structure choice have welfare implications. A marginal increase in q_t means a more efficient risk choice, i.e. a higher expected return, and hence should be welfare improving, *ceteris paribus*. At the same time a marginal increase in k_t implies, due to the equity premium, a higher markup in the intermediation process, which distorts the consumption savings choice and hence lowers welfare, *ceteris paribus*. Since both q_t and k_t are increasing functions of the real interest rate, this begs the question as to whether an increase in the real rate alleviates or intensifies the misallocation due to the banking friction.²⁶ The answer to this question depends on the full set of general-equilibrium conditions. Given the estimated model, we will later numerically verify that the positive first effect dominates, i.e. an increase in R_t^r has welfare improving consequences for the banking market.²⁷ The existence of these opposing welfare effects motivates our optimal policy experiments in section 4.

Finally, the last statement of the proposition implies that a mean-preserving increase in the variance of the real interest rate decreases the mean of the expected return on the bank's investment. This has implications for optimal monetary policy. As we discuss in detail later, the monetary authority cannot affect the nonstochastic steady state of the real rate, but it can influence its volatility. The policy maker therefore has an incentive

²⁵This is true under certainty equivalence, i.e. up to first-order approximation.

²⁶These two opposing forces are well known from the literature on bank capital regulation, where a rise in capital requirements hampers efficient intermediation but leads to a more stable banking sector.

²⁷The dominance of the risk-taking effect is intuitive for two reasons. First, while risk taking entails a real cost, the equity premium just entails a wedge but no direct real costs. Second, as the real interest rate increases the equity premium becomes less important, so a more efficient allocation is intuitive.

to keep the real interest rate stable, at least as long as the opposing effect of the equity premium is negligible.

2.4.5 Full model with deposit insurance and liquidation value

The simplified version of the bank's problem presented so far is useful to explain the basic mechanism. Yet deposit insurance and a non-zero liquidation value are important to improve the quantitative fit of our model to the data.

The assumptions made about deposit insurance and the liquidation value imply that depositors get the maximum of the amount covered by deposit insurance and the value of the capital recovered from a failed project. That means that their return in case of default is:

$$\min \left(\frac{r_{d,t}}{\pi_{t+1}}, \max \left(\frac{r_{k,t+1} + (1-\delta)Q_{t+1}}{Q_t(1-k_t)} \frac{\theta}{1-k_t}, \frac{\psi}{1-k_t} \right) \right).$$

To make deposit insurance meaningful we assume that the liquidation value θ is small enough such that $\frac{r_{k,t+1} + (1-\delta)Q_{t+1}}{Q_t(1-k_t)} \frac{\theta}{1-k_t} < \frac{\psi}{1-k_t}$, which eliminates the inner maximum.²⁸ As the following lemma, proven in Appendix C, states, the outer maximum is unambiguous in equilibrium.²⁹

Lemma: There can be no equilibrium such that the insurance cap is not binding, i.e. $\frac{r_{d,t}}{\pi_{t+1}} > \frac{\psi}{1-k_t}$.

Deposits therefore pay $\frac{\psi}{1-k_t}$ in case of default. Combining the nominal return on the deposit funds (3) with the household's no-arbitrage condition, and defining $\tilde{\psi}_t = E[\Lambda_{t+1}] \psi$, we can write the deposit supply schedule as:

$$q_t \tilde{r}_{d,t} + (1-q_t) \frac{\tilde{\psi}_t}{1-k_t} = \tilde{R}_t. \quad (19)$$

We assume that the deposit insurance scheme, which covers the gap between the insurance cap and the liquidation value for the depositors of failing banks, is financed through a variable tax on capital that is set ex post each period such that the insurance scheme breaks even. The return on loans $\tilde{r}_{l,t}$ can then be rewritten as:

$$\tilde{r}_{l,t} \equiv E_t \left[\Lambda_{t+1} \frac{r_{k,t+1} + (1-\delta)Q_{t+1} - \tau_{t+1}}{Q_t} \right] \quad \text{where} \quad \tau_t = \frac{Q_{t-1} \frac{1-q_{t-1}}{q_{t-1}} \left(\psi - \theta \frac{r_{k,t} + (1-\delta)Q_t}{Q_{t-1}} \right)}{\omega_1 - \frac{\omega_2}{2} q_{t-1}}$$

²⁸In principle the fact that the return on capital is determined only one period later implies that we could have cases where this inequality is satisfied for some states of the world and violated for others. We abstract from this complication, since we later approximate our model locally around the steady state, which allows us to consider only small shocks.

²⁹For this result we again abstract from the effect of uncertainty. See the previous footnote.

This way, the tax also perfectly offsets the distortion in the quantity of investment caused by the deposit insurance. Deposit insurance therefore influences only the funding decision of the bank and, through that, the risk choice. Hence, if q_t was chosen optimally (or was simply a parameter) the deposit insurance would not have any effect.

The same procedure as outlined above can be applied to obtain closed-form solutions³⁰ for the risk choice and the equity ratio. The solutions can be found in Appendix C. As stated below in proposition 2, the equilibrium characterizations in subsection 2.4.4 remain valid. In particular, note that the deviation of the chosen risk (equity ratio) from the optimal level decreases (increases) in the real interest rate. Given our estimation, the risk effect dominates in terms of welfare implications. The intuition for the risk-taking channel is similar to before.

Deposit insurance makes deposits cheaper relative to equity. As a result, the bank demands more deposits and chooses a riskier investment portfolio. Deposit insurance furthermore strengthens the risk-taking channel, which is now affected not only by the importance of the equity premium relative to the real interest rate, but also by the importance of the deposit insurance cap relative to the real interest rate. On the other hand, the efficient risk level is not affected by deposit insurance.

The liquidation value, on the other hand, is irrelevant for the banks' and investors' choice, since it is assumed to be smaller than deposit insurance. Yet it eases the excessiveness of risk taking, since it increases the optimal level of risk: $q_t^o = \frac{\omega_1 - \theta}{\omega_2}$.

An additional implication of our model, given the estimated parameters obtained in section 3, is that both the expected loan profitability and the risk premium fall as the real interest rate declines: a result which is in line with the respective findings of Ioannidou et al. (2014) and Buch et al. (2014).

Finally, we would like to point out that none of the results in proposition 1 is due to the functional form that we have assumed for the risk return trade-off. The statement holds even for a generic function $f(q_t)$ ³¹ under relatively weak assumptions, some of which are sufficient but non necessary. For a proof and a discussion of these assumptions see Appendix C.

Proposition 2: Consider proposition 1, but replace $f(q_t)$ with the expected return, taking into account the liquidation value of failed projects: $f(q_t) + (1 - q_t)\theta$.

- (1) Given this adjustment, all statements of proposition 1 hold for the full bank model with deposit insurance and a small enough liquidation value as well.
- (2) Given this adjustment, statements (1)-(4) of proposition 1 hold for a generic conditional expected return function $f(q_t)$ with deposit insurance and a small

³⁰In this case, one needs to apply the adjusted deposit supply schedule (19) and to make sensible assumptions about the relative size of parameters and about the root when solving the zero-profit equation.

³¹Given that the recovery value $f(q_t)$ now describes the expected return *conditional on success*.

enough liquidation value under the additional assumptions that $f(q_t)$ satisfies $f(q_t) \geq 0$, $f''(q_t) < 0$, $f'''(q_t) \leq 0$, $f''''(q_t) \leq 0$. Statement (5) holds if furthermore either the default probability is low relative to the parameters $\frac{q_t}{(1-q_t)}\tilde{\xi}_t \geq \tilde{R}_t - \tilde{\psi}_t$ or there is no deposit insurance $\tilde{\psi}_t = 0$.

2.5 Labour and goods sectors

The labour and goods sectors feature monopolistic competition and nominal rigidities as Calvo (1983), which allow for a role for monetary policy. Since the modeling of these sectors follows the canonical New Keynesian model, we discuss them briefly in Appendix B and refer to Smets and Wouters (2007) and Adjemian et al. (2008) for further details.

2.6 Monetary and fiscal policy

The central bank follows a nominal interest rate rule, targeting inflation and output deviations from the steady state. In addition, the fiscal authority finances a stochastic expenditure stream $g_y \bar{Y} \varepsilon_t^G$ through lump sum taxes.

3 Dynamic implications of the risk-taking channel in the estimated model

We have embedded our risk-taking channel in a medium-scale model which closely resembles the non-linear version³² of Smets and Wouters (2007), and we next estimate the model parameters using Bayesian techniques. This serves two purposes. First, we want to assess whether the risk-taking channel improves the quantitative fit of the model, once other monetary and real frictions are taken into account. Second, Smets and Wouters (2007) provide a quantitative model that is able to replicate key empirical moments of the data, which are needed for the monetary policy evaluation we perform in Section 4.

3.1 Model estimation

We estimate a linearized version of the model with Bayesian techniques using eight US macroeconomic time series covering the period of the great moderation from 1984Q1 to 2007Q3. These include the seven series used by Smets and Wouters (2007), i.e. the federal funds rate, the log of hours worked, inflation and the growth rates in the real hourly wage and in per-capita real GDP, real consumption, and real investment. To identify the banking sector parameters we add a series of the banking sector equity ratio,

³²Our model deviates from Smets and Wouters (2007) only to the extent that we abstract (for simplicity) from capital utilization, shown by the authors to be of secondary importance once wage stickiness is taken into account, and growth. Furthermore, since we use one additional time series we have added a time preference shock and reinterpreted it as an equity premium shock that affects only bank equity.

which we construct from aggregate bank balance-sheet data provided by the FDIC. For a full description of the data we refer to Appendix A and to the supplementary material of Smets and Wouters (2007). The observation equations, linking the observed time series to the variables in the model, as well as the prior specifications can be found in Appendix B. While the priors of the non-bank parameters follow Smets and Wouters (2007), the priors for the banking sector parameters are motivated by historical averages and external estimates for the US. Note that instead of forming priors directly about ω_2 (risk return trade-off) and ψ (deposit insurance), we rewrite these parameters as functions of the steady-state equity ratio \bar{k} and default rate \bar{q} . The prior mean of the steady-state equity premium ξ is centred around an annualized value of 6%, in line with the empirical estimates of Mehra and Prescott (1985), while the prior distribution for \bar{k} is diffuse and centred around the historic mean of 12%. The prior for the liquidation value θ is set such that it is contained between 0.3 and 0.7 with a 95% probability, in line with the evidence provided by Altman et al. (2003). The success rate \bar{q} is not well identified and is therefore fixed to 0.99, which implies an annual default rate of 4%, roughly in line with the historical average of delinquency rates on US business loans. Sensitivity tests have moreover shown that this parameter is only of small quantitative relevance.³³ Lastly, we normalize the units of capital versus final goods by setting ω_1 (return on the risky asset) such that one unit of final good is expected to produce one unit of capital good in steady state.

Table 4 in Appendix B summarizes the posterior parameter values, which are broadly in line with existing empirical estimates for the US. The key banking sector parameters that determine the importance of the risk-taking channel are well identified by the data. The steady-state equity ratio has a tight posterior around 12%, the posterior mean of the equity premium is around an annualized value of 9%, and the liquidation value is about 74%.³⁴ For the following quantitative analysis we set the parameters to their posterior means.

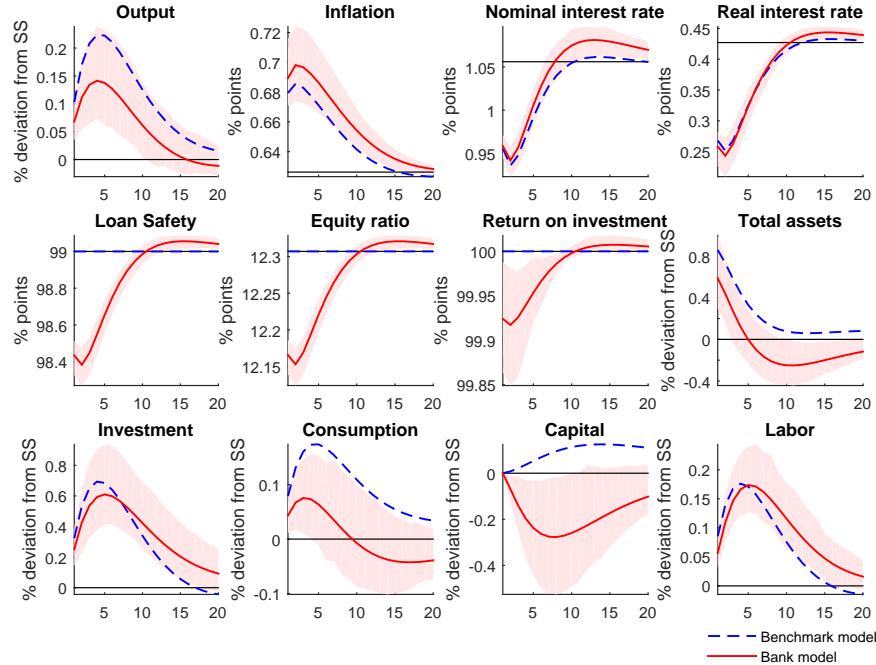
3.2 Dynamic implications of excessive risk taking

To illustrate the dynamic effects of the risk-taking channel, we assess how the propagation of monetary policy shocks is affected by the risk-taking channel. For this purpose, we compare the impulse responses of two models: with banking frictions (henceforth bank model), and without banking frictions. To enable the dynamic comparison, we equalize

³³In particular, the implications for optimal monetary policy behaviour are very robust to the value of the steady state default rate. What matters is the importance of the channel over the business cycle, determined predominantly by the liquidation value and the scope of deposit insurance.

³⁴The implied mean value for deposit insurance cap ψ of about 88% implies that 99% of deposits are insured in steady state. Demirgüç-Kunt et al. (2005) report that the explicit deposit insurance scheme in the US is estimated to cover between 60% and 65% of deposits. The divergence can be interpreted as implicit deposit guarantees resulting from the expectation of bailouts. The implied mean values of ω_1 (1.13) and ω_2 (0.2561) yield a corner solution for q^{opt} at 1.

Figure 1: **Monetary policy shock in the bank and benchmark models:** dynamic responses in the bank model (solid red lines) and in the benchmark model (dashed blue lines) to an expansionary monetary policy shock, at the mean of the posterior distribution. Shaded areas denote the highest posterior density interval at 90% for the bank model impulse responses, and the black line the steady-state level. Inflation and interest rates are quarter-on-quarter rates.



the steady states of the two economies. In particular, we alter the model without financial frictions by treating the risk choice q_t and the equity ratio k_t as fixed parameters, which we set to the steady-state values of the bank model. This model, henceforth benchmark model, has the same steady state as the bank model and corresponds to a standard New Keynesian model with a small markup in capital markets.

In Figure 1 we compare the dynamic responses in the bank model (solid red lines) and in the benchmark model (dashed blue lines) to an expansionary monetary policy shock. A monetary policy expansion triggers a set of standard reactions, which are evident in the benchmark model. An unexpected fall in the nominal risk-free rate causes a drop in the real interest rate, since prices are sticky. Consequently, consumption is shifted forward, firms that can adjust the price do so, causing an increase in inflation, while the remaining firms increase production. The risk-taking channel adds two further elements, as both the risk level and the capital structure chosen by the bank respond to the real interest rate movement. On impact, the drop in the real interest rate cause banks to substitute equity for deposits, since the relative cost advantage of deposits increases. Consequently, banks have less skin in the game and hence take more risk (lower loan safety). The risk choice therefore moves further away from the optimal level, and the expected return on aggregate investment $f(q_t) = q_t(\omega_1 - \omega_2/2q_t) + (1 - q_t)\theta$ drops.³⁵ To maintain the same path of

³⁵Note that the decline in the equity ratio diminishes the distortion due to the equity premium, which reduces the cost of capital. Yet this effect is tiny relative to the increase in the cost of capital due to lower investment efficiency.

capital as in the benchmark case, households would have to invest more and consume less. Yet this would not be optimal because of consumption smoothing and because of the lower expected return on investment. Therefore, investment rises by less than what would be needed to compensate the loss in investment efficiency, which makes the capital stock decline considerably. Overall, agents are worse off (in terms of welfare) in the bank model than in the benchmark economy.

3.3 Evaluating the fit of the estimated model

Comparing our bank model to the benchmark Smets and Wouters (2007)-type model, we find that the fit of the bank model is better.³⁶ In particular, the posterior odds ratio of $\exp(2.86)$ can be interpreted as providing 'positive' evidence in favour of the bank model, according to Jeffreys (1961) and Kass and Raftery (1995). To evaluate the plausibility of the *strength* of the proposed channel, we furthermore estimate a meta-model that nests both the benchmark and the bank model. This meta-model is characterized by the auxiliary parameter Γ , which defines the weight on each of the two models.³⁷ Using a flat prior we find that the posterior mean lies around -0.5, implying that the benchmark model is rejected and the estimated risk-taking channel is slightly too weak, if anything.

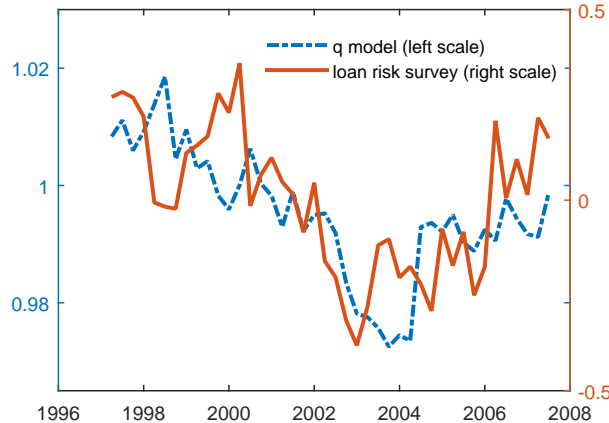
A close examination of the role of the investment efficiency shock in the two estimated models provides some intuition for why the risk-taking channel improves the fit of the model. In particular, we find that the introduction of the banking frictions reduces the forecast error variance of output by a third, while the variance decomposition share of the investment shock drops from around 49 percent (estimated benchmark model) to 34 percent (estimated bank model) for horizons between 3 and 8 quarters. This relates to the argument of Justiniano et al. (2011), who find that the major role of this shock in explaining GDP volatility in the canonical medium-scaled Smets and Wouters (2007) model could be a spurious result that captures unmodelled financial frictions. In reducing the importance of this shock, the risk-taking channel seems to be capable of capturing at least some of this missing mechanism. This is intuitive because both the investment shock ε_t^I and the expected return on the banks' investment $q_t \left(\omega_1 - \frac{\omega_2}{2} q_t \right) + (1 - q_t)\theta$ enter the capital accumulation equation multiplicatively:

$$K_t = \left[\varepsilon_t^I (1 - S(i_t/i_{t-1})) i_t + (1 - \delta) K_{t-1} \right] \left[q_t \left(\omega_1 - \frac{\omega_2}{2} q_t \right) + (1 - q_t)\theta \right].$$

³⁶Recall that the Smets and Wouters (2007)-type model is obtained by turning off the banking sector frictions. Hence bank leverage is no longer defined. For the comparison we therefore estimate the two versions of the model (with and without the banking frictions) using only the seven macro aggregates used by Smets and Wouters (2007), and calibrate the banking parameters in the bank model to the posterior estimates in Table 4.

³⁷To set up this meta-model, we replace the equations for k_t and q_t in the bank model with: $k_t = \Gamma k_t^{benchmark} + (1 - \Gamma) k_t^{bank}$ and $q_t = \Gamma q_t^{benchmark} (1 - \Gamma) q_t^{bank}$, where k_t^x corresponds to the RHS of the corresponding equation from the x model. Again, this estimation is run using the 7 non-financial time series with fixed banking sector parameters.

Figure 2: **Risk taking in the model and in the data:** The figure compares the value of loan safety q_t implied by the estimated model (in particular we plot the mean of the series posterior distribution) with a survey-based index of loan safety computed from the US Terms of Business Lending Survey. For a more detailed discussion see text and Appendix D.



Yet the two are not perfectly isomorphic, since the shock affects only net investment, while the expected return on investment affects gross investment. Moreover, the path of ε_t^I backed out from the estimated benchmark model is strongly correlated with the path of the return on investment in the estimated bank model.³⁸

To evaluate the fit with respect to financial variables we look at two statistics that were not targeted by the estimation. First, we compare the model-implied series for the risk variable q_t with a survey-based proxy for bank risk taking. The latter is a weighted average³⁹ of the internal risk rating assigned by banks to newly issued loans, provided by the US Terms of Business Lending Survey, and inverted so as to match the definition of q_t in the model. Figure 2 shows that the model implies a cyclical pattern of risk that is roughly in line with the survey measure (the correlation is 60 percent). Second, the responses in Figure 1 also show that, conditional on the monetary policy shock, leverage (the inverse of the equity ratio k_t) is pro-cyclical with respect to the size of the bank balance sheet $e_t + d_t$. Conditional on the full set of shocks, we find a correlation of 43 percent which is in line with the evidence for US data provided by Adrian and Shin (2014), and distinguishes our model from canonical financial accelerator models that build on Bernanke et al. (1999).⁴⁰ Note in addition that the responses shown in figures 1 are in accordance with the structural VAR results in appendix D, in particular with the finding that the response of risk is proportional to that of the interest rate.

Overall, these findings suggest that the inclusion of the risk-taking channel improves

³⁸For this exercise we use the same specification as for the likelihood comparison. Notice that the specification of our model is not exactly the same as Smets and Wouters (2007) and Justiniano et al. (2011) since we have abstracted from capital utilization and use internal rather than external habits. This means that the numbers are not directly comparable.

³⁹For a detailed discussion about this variable we refer to Appendix D.

⁴⁰See, for instance, the discussion in Adrian et al. (2015).

the fit of nonfinancial data, while at the same time matching two key characteristics of aggregate banking sector data.

4 Monetary policy with the risk-taking channel

Optimal monetary policy in the benchmark Smets and Wouters (2007)-type model is well understood. Optimal monetary policy in this economy generally aims at mitigating the effect of coexisting nominal frictions (price and wage stickiness) and real frictions (monopolistic competition in goods and labour markets). The effect of price stickiness can be avoided by keeping inflation stable. While wage stickiness and the real frictions provide a motive to deviate from strict inflation stabilization, their effect is typically quantitatively small and it is therefore, as Woodford (2004) puts it, “not a bad first approximation to say that the goal of monetary policy should be price stability”. This goal is easy to implement with a simple Taylor rule that attaches a major weight on inflation.

The financial sector modelled in this paper adds another real friction: inefficient risk taking. As we have seen, the intensity of this distortion depends on the real interest rate. Upward movements in the real interest rate increase the efficiency of the capital production technology since they lower the level of risk taking towards the efficient level. Downward changes in the real interest rate lead instead to even more excessive risk taking. But this does not mean that movements in the real rate are on average irrelevant. We have shown in section 2 that the expected return on investments $f(q_t)$ is concave in the real interest rate. A mean-preserving spread of the real rate therefore reduces the average expected return on investments. Therefore the risk-taking channel provides a motive for keeping the real interest rate constant. Yet this objective conflicts with the inflation stabilization objective resulting from nominal rigidities, since the latter call for aggressive movements in both the nominal and the real rate. The risk-taking channel therefore introduces a new monetary policy trade-off. As a result, the central bank should accept higher inflation volatility in order to reduce the distortion stemming from risk taking.

But is this new trade-off actually quantitatively significant for monetary policy? Given that many previous studies have found the optimality of inflation stabilization to be robust to the introduction of financial frictions (e.g. De Fiore and Tristani (2013) and Bernanke and Gertler (2001)), the answer to this question is not trivial. To answer this question we determine *optimal simple implementable* monetary policy rules in both the bank and benchmark models, and compare their performance in the bank economy.⁴¹ This comparison has an interesting interpretation. Suppose that the actual economy features

⁴¹Note that for this experiment we use the parameters estimated for the bank model both for the bank and benchmark economy.

the risk-taking channel (the bank model), but that the central bank is unaware of this channel and believes that risk cannot be influenced by the interest rate. The central bank would then implement optimal policy based on a wrong model (the benchmark model). Our comparison then answers the question of how important it is to understand the risk-taking channel, in terms of optimal policy and welfare.

Notice that in this paper we consider a central bank that has no policy tools besides the interest rate.⁴² Exploring optimal macroprudential regulation is however beyond the scope of the present paper.⁴³

4.1 The central bank problem

We follow Schmitt-Grohe and Uribe (2007) and characterize optimal monetary policy as the policy rule that maximizes welfare among the class of simple, implementable interest-rate feedback rules⁴⁴ given by:

$$R_t - \bar{R} = \phi_\pi \hat{\pi}_{t+s} + \phi_y \hat{y}_{t+s} + \phi_k \hat{k}_{t+s} + \rho (R_{t-1} - \bar{R}) , \quad (20)$$

where the hat symbol denotes (expected) percentage deviations from the steady state, and the index s allows for forward- or contemporaneous-looking rules (respectively by setting $s = 1$ or $s = 0$). The policy rule specification (20) is chosen for its generality, as it encompasses both standard Taylor-type rules (setting $\phi_k = 0$), and the possibility that the central bank reacts to banking sector leverage, the inverse of the equity ratio k ($\phi_k \neq 0$). A fall in the equity ratio implies that banks increase their debt financing, i.e. they increase leverage. As a consequence, banks internalize less the downside risk of their investments, and choose loans with a higher default probability. Hence, a fall in the equity ratio signals an increase in risk taking, to which the central bank may want to respond by increasing the interest rate. We choose not to let the interest rate depend on risk taking directly, because the latter is not a readily observable variable. We furthermore impose that the inertia parameter ρ has to be non-negative. Since we are interested in the effect of systematic monetary policy, we switch off the monetary policy shock for this experiment.

The welfare criterion that defines the optimal parameter combination for rule (20) is

⁴²In the model, with a second instrument, such as capital regulation, the central bank could do better or even eliminate the friction. In practice, however, regulation has not been able to eliminate the risk-taking channel in the past. While improvements in regulatory practice are certainly feasible, we are sceptical whether it will be possible to *fully* eliminate the channel through regulation. It is therefore likely that the channel will continue to play a role for monetary policy.

⁴³For a thorough analysis of macroprudential policy in an economy with bank risk-taking see Collard et al. (2012).

⁴⁴The implementability criterion requires uniqueness of the rational expectations equilibrium, while simplicity requires the interest rate to be a function of readily observable variables. For a complete discussion, see Schmitt-Grohe and Uribe (2007). Notice that we drop their second requirement for implementability, which is that an implementable rule must avoid regular zero lower bound violations.

Table 1: **Optimal simple rules:** The first (second) column describes the timing (restrictions) of the policy rule. The last row corresponds to a rule where only ϕ_k can be chosen and the other parameters are fixed to the optimized benchmark values. V is the welfare level associated with each policy in the bank model. Ω is the welfare cost (in % of the consumption stream) associated with implementing in the bank model the optimal policy rule of the benchmark model. For the benchmark model the restriction $\phi_k = 0$ is irrelevant, since the equity ratio is constant. Italics indicate restricted parameters.

s	rule	<i>benchmark model</i>			<i>bank model</i>					
		ρ	$\phi_{\pi_{t+s}}$	$\phi_{y_{t+s}}$	ρ	$\phi_{\pi_{t+s}}$	$\phi_{y_{t+s}}$	ϕ_{k_t}	V	Ω
0	$\phi_k, \rho = 0$	<i>0</i>	7.100	0.115	<i>0</i>	3.080	0.126	<i>0</i>	-185.321	0.476
0	$\phi_k = 0$	0.000	7.100	0.115	1.059	0.510	0.005	<i>0</i>	-184.750	0.898
0	$\rho = 0$	<i>0</i>	7.100	0.115	<i>0</i>	2.637	0.097	0.027	-185.314	0.481
1	$\phi_k, \rho = 0$	<i>0</i>	17.222	0.148	<i>0</i>	4.294	0.172	<i>0</i>	-185.209	0.687
1	$\phi_k = 0$	0.236	12.084	0.124	1.114	0.072	0.074	<i>0</i>	-184.656	0.813
0	choose ϕ_k	0.000	7.100	0.115	<i>0</i>	<i>7.100</i>	<i>0.115</i>	-0.177	-185.438	0.389

the household's conditional lifetime utility.⁴⁵ In order to compare welfare levels we define the measure Ω as the fraction of the consumption stream that a household would need to receive as a transfer under the suboptimal rule to be equally well off as under the optimal rule. If o denotes the optimal rule and s another suboptimal rule, this fraction Ω is implicitly defined by the equation:

$$V^o = E_0 \sum_{t=0}^{\infty} \beta^t \varepsilon_t^B u((1 + \Omega)c_t^s, L_t^s) .$$

4.2 Findings

Using the welfare criterion just described, we numerically determine the coefficients of the optimal simple implementable rules in the benchmark and in the bank model, using second-order approximations around the non-stochastic steady state. The first five rows of Table 1 report the optimal coefficients for five different specifications of the monetary policy rule: contemporaneous and forward-looking, without inertia and with optimal inertia, without and with a reaction to current leverage. The coefficients of the optimal rules generally vary greatly between the two models. A set of results which are robust across policy rule and estimation⁴⁶ specifications, are worth noticing.

First, the optimal coefficients on inflation deviations are smaller in the bank model compared to the benchmark model. Given that the optimal output coefficient is close to zero, the optimal rule is hence closer to a stable real interest rate rule in the bank model than in the benchmark model. Furthermore, if the central bank can optimize over its smoothing parameter, then full interest rate smoothing is optimal in the bank model.

⁴⁵ This measure, which is conditional on the economy being in steady state, is common in the literature.

⁴⁶ We have experimented with different estimation samples and calibrated parameter values. While the optimized parameters and transfers slightly change, the qualitative results discussed in the text are very robust.

Table 2: **Differences in moments associated with the optimal simple rules in the benchmark and in the bank model:** This Table shows the % differences in the mean and standard deviation associated with applying the different optimal rules in the bank model. The first entry, for example, indicates that under the optimal bank policy rule average risk would be 0.15% lower than if the rule optimal for the benchmark model had been applied.

<i>s</i>	rule	<i>mean</i>						<i>standard deviation</i>				
		<i>q</i>	R^r	π	<i>y</i>	<i>c</i>	$f(q)$	<i>q</i>	R^r	π	<i>y</i>	<i>c</i>
0	$\phi_k, \rho = 0$	0.151	0.002	-0.051	0.311	0.499	0.045	-43.880	-47.975	52.470	-0.843	-4.108
0	$\phi_k = 0$	0.214	0.007	-0.038	0.439	0.701	0.061	-67.949	-77.760	64.393	-9.545	-9.566
0	$\rho = 0$	0.152	0.003	-0.015	0.323	0.506	0.045	-41.666	-47.248	53.194	-0.773	-3.800
1	$\phi_k, \rho = 0$	0.194	0.011	-0.037	0.413	0.652	0.062	-50.536	-55.417	57.719	-2.781	-6.906
1	$\phi_k = 0$	0.195	0.004	-0.054	0.458	0.724	0.057	-65.839	-76.3112	71.906	-10.373	-11.737
0	choose ϕ_k	0.130	0.001	-0.070	0.244	0.417	0.042	-41.691	-41.948	31.838	3.323	-0.345

Table 2, which displays the changes in the mean and volatility of key variables in switching from the benchmark optimal rule to the bank optimal rule in the bank model, helps to understand the rationale behind these results. By responding less aggressively to inflation and by smoothing the nominal interest rate, the central bank limits fluctuations in the real interest rate. The lower volatility of R_t^r translates into a higher average return on investment $f(q_t)$ due to the concavity of this function in R_t^r .⁴⁷ This higher average return on investment comes at the cost of a significantly higher inflation volatility. Hence, the new trade-off between inflation and real rate stabilization implies a significant deviation from inflation stabilization.

To understand *how different* the equilibria associated to the two optimal rules are, and therefore how important it is for the central bank to take the risk-taking channel into account, we compute the cost Ω of applying the rule that is optimal for the benchmark model in the bank model. These costs, expressed in % of the lifetime consumption stream, are reported in the last column of Table 1. Though the costs vary greatly across policy specifications, they are always significant. For the best-performing policy (fifth row of Table 1), the costs of applying the benchmark policy in the bank model are around 0.81% of the lifetime consumption stream. Unlike the financial frictions analyzed by De Fiore and Tristani (2013) and Bernanke and Gertler (2001), the risk-taking channel therefore has implications for monetary policy that are important both in terms of the prescribed policy and the cost of deviating from it.

Second, including an explicit response to banking sector leverage, in addition to inflation and output, improves welfare only marginally (compare the last column of the first and third row of Table 1). Recall that leverage depends on both the nominal interest rate and expected inflation. By setting the nominal rate optimally as a function of current inflation, the central bank can already steer risk taking, to the extent that current and expected future inflation are highly correlated. The fact that this correlation is not perfect,

⁴⁷Note that the slight increase in R_t^r accounts only for a marginal fraction of the increase in $f(q_t)$.

and that our approximation allows for nonlinearities, accounts for the small improvement in welfare obtained by allowing a response to leverage in the policy function. To further illustrate this point, in the last row we fix the coefficients of current inflation and output to the values which are optimal in the benchmark economy, and allow the central bank to respond optimally only to leverage. In this case, it is optimal to strongly raise the interest rate in response to higher leverage (lower equity ratio k). Thereby the central bank again stabilizes the real interest rate and does not do much worse in terms of welfare than when the responses to inflation and output are chosen optimally (compare the last column of the third and sixth row of Tables 1 and 2).

5 Conclusion

Recent empirical evidence suggests that monetary policy can influence bank risk-taking behaviour. However, the economic relevance of this connection and its implications for monetary policy are still unclear. We address these questions by developing and estimating a quantitative general-equilibrium model where interest rates affect bank risk taking.

In the model, all savings are intermediated by banks, which make risky investments. Low levels of the risk-free interest rate induce banks to make riskier and therefore less efficient investments. At the core of this mechanism is an agency problem between depositors and equity providers: the latter choose the level of risk but are protected by limited liability. The response of risk taking to interest rates alters the dynamics in the model. To validate the risk-taking channel and gauge its importance, we estimate the model with Bayesian methods for the US. We find that this new channel not only improves the fit of the model but also predicts a path for risk taking that is in line with survey evidence.

We use the model to derive both analytical and quantitative implications for monetary policy. We show analytically that the risk-taking channel implies a motive for real interest rate stabilization. Since this motive conflicts with the inflation stabilization objective, it generates a new trade-off for the central bank. We then use the estimated model to analyze the quantitative importance of this new trade-off. Contrary to similar experiments with financial frictions in the literature, we find that the risk-taking channel calls for significant deviations from inflation stabilization. The policy maker optimally accepts 50% more inflation volatility to reduce the volatility of the real rate by a similar share. Taking the risk-taking channel into account generates welfare gains for consumers between 0.5% and 1%.

References

- Adjemian, S., Darracq Paries, M., and Moyen, S. (2008). Towards a monetary policy evaluation framework. Working Paper Series 0942, European Central Bank.
- Adrian, T., Boyarchenko, N., and Shin, H. S. (2015). The cyclicality of leverage. Staff Reports 743, Federal Reserve Bank of New York.
- Adrian, T. and Shin, H. S. (2014). Procyclical Leverage and Value-at-Risk. *Review of Financial Studies*, 27(2):373–403.
- Afanasyeva, E. and Guentner, J. (2015). Lending Standards, Credit Booms, and Monetary Policy. Technical report, mimeo.
- Altman, E., Resti, A., and Sironi, A. (2003). Default Recovery Rates in Credit Risk Modeling: A Review of the Literature and Empirical Evidence. Working Paper Series S-CDM-03-11, NYU Working Paper.
- Angeloni, I. and Faia, E. (2013). Capital regulation and monetary policy with fragile banks. *Journal of Monetary Economics*, 60(3):311–324.
- Angeloni, I., Faia, E., and Lo Duca, M. (2015). Monetary policy and risk taking. *Journal of Economic Dynamics and Control*, 52:285–307.
- Bernanke, B. S. and Gertler, M. (2001). Should Central Banks Respond to Movements in Asset Prices? *American Economic Review*, 91(2):253–257.
- Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In Taylor, J. B. and Woodford, M. D., editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 21, pages 1341–1393. Elsevier.
- Borio, C. and Zhu, H. (2008). Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism? (268).
- Buch, C. M., Eickmeier, S., and Prieto, E. (2014). In search for yield? Survey-based evidence on bank risk taking. *Journal of Economic Dynamics and Control*, 43(C):12–30.
- Calvo, G. A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics*, 12(3):383–398.
- Christensen, I., Meh, C., and Moran, K. (2011). Bank Leverage Regulation and Macroeconomic Dynamics. Working Papers 11-32, Bank of Canada.

- Collard, F., Dellas, H., Diba, B., and Loisel, O. (2012). Optimal Monetary and Prudential Policies. Technical report.
- De Fiore, F. and Tristani, O. (2013). Optimal Monetary Policy in a Model of the Credit Channel. *Economic Journal*, 123(571):906–931.
- de Groot, O. (2014). The Risk Channel of Monetary Policy. *International Journal of Central Banking*, 10(2):115–160.
- Dell’Ariccia, G., Laeven, L., and Marquez, R. (2014). Real interest rates, leverage, and bank risk-taking. *Journal of Economic Theory*, 149(0):65 – 99.
- Dell’Ariccia, G., Laeven, L., and Suarez, G. (2013). Bank Leverage and Monetary Policy’s Risk-Taking Channel: Evidence from the United States. Working Paper 13/143, IMF.
- Demirgüç-Kunt, A., Karacaovali, B., and Laeven, L. (2005). Deposit Insurance Around the World: A Comprehensive Database. World Bank Policy Research Working Paper 3628.
- Gertler, M., Kiyotaki, N., and Queralto, A. (2012). Financial crises, bank risk exposure and government financial policy. *Journal of Monetary Economics*, 59(S):S17–S34.
- Ioannidou, V., Ongena, S., and Peydro, J.-L. (2014). Monetary Policy, Risk-Taking and Pricing: Evidence from a Quasi-Natural Experiment. Discussion Paper 2009-04S, European Banking Center.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, 3rd edition.
- Jimenez, G., Ongena, S., Peydro, J. L., and Saurina, J. (2014). Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking? *Econometrica*, 82(2):463–505.
- Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2011). Investment shocks and the relative price of investment. *Review of Economic Dynamics*, 14(1):102–121.
- Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kimball, M. S. T. (1995). The Quantitative Analytics of the Basic Neomonetarist Model. *Journal of Money, Credit and Banking*, 27(4):1241–77.
- Mehra, R. and Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161.
- Schmitt-Grohe, S. and Uribe, M. (2007). Optimal, Simple, and Implementable Monetary and Fiscal Rules . *Journal of Monetary Economics*, 54(6):1702–1725.

- Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.
- Taylor, J. B. (2007). Housing and monetary policy. *Proceedings - Economic Policy Symposium - Jackson Hole*, pages 463–476.
- Woodford, M. (2004). Inflation Targeting and Optimal Monetary Policy . *The Federal Reserve Bank of St. Louis Review*, 90(430):15–42.

Appendices

Appendix A: Data description

Table 3: All level variables are expressed in per-capita terms (divided by N). Hours are measured as $H_1 \cdot H_2/N$ where H_1 is converted into an index. The nominal wage W is deflated by the GDP deflator. We define equity capital as equity plus reserves plus subordinated debt, and total liabilities as equity plus deposits. To do so, we net out two types of liabilities, since they are typically overcollateralized: federal funds purchased & repurchase agreements and federal home loan bank advances. Furthermore, we omit a few categories of debt that match neither of our concepts of insured deposits and equity, or that are simply not enough characterized: other borrowed money, uncategorized liabilities, trading book liabilities, banks liability on acceptances. All of these balance sheet positions are minor. Over the observation period, the first group accounts for roughly 11% of the balance sheet, the second for about 9%. All indexes are adjusted such that 2009 = 100. The estimation sample spans from 1984Q1 to 2007Q3. Our survey-based proxy for bank risk-taking q , is constructed using data from the US Terms of Business Lending Survey. The survey provides the internal risk rating assigned by banks to newly issued loans. In this survey, available from 1997Q1 onwards, 400 banks report the volume of loans originated in the first week of the mid month each quarter, grouped by internal risk rating. This rating varies between 1 and 5, with 5 being the maximum level of risk. Following Dell’Ariccia et al. (2014), we construct a weighted average loan risk series, using as weights the value of loans in each risk category.

SYMBOL	SERIES	MNEMONIC	UNIT	SOURCE
Y	REAL GROSS DOMESTIC PRODUCT	GDPC96	BN. USD	FRED / BEA
P	GDP DEFLATOR	GDPDEF	INDEX	FRED / BEA
R	EFFECTIVE FEDERAL FUNDS RATE	FEDFUNDS	%	FRED / BOARD OF GOVERNORS
C	PERSONAL CONSUMPTION EXPENDITURE	PCEC	BN. USD	FRED / BEA
I	FIXED PRIVATE INVESTMENT	FPI	BN. USD	FRED / BEA
H_1	CIVILIAN EMPLOYMENT	CE16OV	THOUSANDS	FRED / BLS
H_2	NONFARM BUSINESS (..) HOURS	PRS85006023	INDEX	DEPARTMENT OF LABOR
W	NONFARM BUSINESS (..) HOURLY COMPENSATION	PRS85006103	INDEX	DEPARTMENT OF LABOR
N	CIVILIAN POPULATION	LNS1000000	0CE16OV	BLS
q	AVERAGE WEIGHTED LOAN RISK	own calculation	%	BOARD OF GOVERNORS
E	EQUITY CAPITAL OVER LIABILITIES	own calculation	%	FDIC

Appendix B: The full model: equilibrium and estimation details

This Appendix outlines the set-up of the conventional parts of the model. Furthermore, it defines the equilibrium, listing all model equations, grouped by sector.

The Households' problem: Households maximize their lifetime utility function:

$$\max_{d_t, e_t, s_t, c_t, L_t} E \left[\sum_{t=0}^{\infty} \beta^t \varepsilon_t^B \frac{(c_t - \iota c_{t-1})^{1-\sigma_C}}{1-\sigma_C} \exp \left(\varphi L_t^{1+\sigma_L} \frac{\sigma_C - 1}{1 + \sigma_L} \right) \right] ,$$

subject to the per-period budget constraint in real terms:

$$c_t + d_t + e_t + s_t + T_t = L_t w_t + d_{t-1} \frac{R_{d,t}}{\pi_t} + e_{t-1} \frac{R_{e,t}}{\pi_t} + s_{t-1} \frac{R_{t-1}}{\pi_t} + \Pi_t ,$$

where π_t is the inflation rate, while T_t and Π_t are taxes and profits from firm ownership, expressed in real terms. We allow for habits in consumption (ι) and a time preference shock ε_t^B . This shock is assumed to be persistent with log-normal innovations, like all following shocks unless otherwise specified. The household's optimality conditions are given by the usual Euler equation, two no-arbitrage conditions and the labour supply condition:

$$\begin{aligned} \Lambda_t &= \beta E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] , \\ E_t \left[\Lambda_{t+1} \frac{R_{d,t+1}}{\pi_{t+1}} \right] &= E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] , \\ E_t \left[\Lambda_{t+1} \frac{R_{e,t+1}}{\pi_{t+1}} \right] &= E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] , \\ \Lambda_t w_t &= \frac{(c_t - \iota c_{t-1})^{1-\sigma_C}}{1-\sigma_C} \exp \left(\varphi L_t^{1+\sigma_L} \frac{\sigma_C - 1}{1 + \sigma_L} \right) (1 - \sigma_C) L_t^{\sigma_L} \end{aligned}$$

where $\Lambda_t = \varepsilon_t^B (c_t - \iota c_{t-1})^{-\sigma_C} - \beta \iota E_t [\varepsilon_{t+1}^B (c_{t+1} - \iota c_t)^{-\sigma_C}]$ is the marginal utility of consumption.

Labour and goods sectors: Final goods producers assemble different varieties of intermediate goods through a Kimball (1995) aggregator with elasticity of substitution ϵ_p and Kimball parameter k_p , taking as given both the final good price and the prices of intermediate goods. Their optimization problem yields demand functions for each intermediate good variety as a function of its relative price.

A continuum of firms produces differentiated intermediate goods using capital K_{t-1} and “packed” labour l_t^d as inputs. The production function is Cobb-Douglas and is affected by a total factor productivity shock ε_t^A . Firms use their monopolistic power to set prices, taking as given their demand schedule. As in Calvo (1983), they can reset their prices in

each period with probability λ^p , otherwise they index their prices to past inflation with degree γ^p and to steady-state inflation with degree $(1 - \gamma^p)$. Furthermore, they are subject to a time-varying revenue tax ε_t^p that is equivalent to a markup shock, up to a first-order approximation.

The labour market resembles the product market. Packed labor is produced by labor packers, who aggregate differentiated labor services using a Kimball (1995) aggregator with elasticity of substitution ϵ_w and Kimball parameter k_w .

Differentiated labour services are produced by a continuum of unions from the households labour supply. They use their monopolistic power to set wages. Wages are reset with probability λ^w , otherwise they are indexed to past inflation (with degree γ^w) and steady-state inflation. Like intermediate firms, unions are subject to a stochastic wage tax ε_t^w .⁴⁸

See the Appendix of Smets and Wouters (2007) for a detailed discussion of the set-up and Adjemian et al. (2008) for the recursive formulation of the equilibrium conditions, which we list below.

Monetary and fiscal policy: The central bank follows a nominal interest rate rule, targeting inflation and output deviations from the steady state:

$$R_t - \bar{R} = (1 - \rho) (\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t) + \rho (R_{t-1} - \bar{R}) + \varepsilon_t^R, \quad (21)$$

where ρ is a smoothing parameter, the hat symbol denotes percentage deviations from the steady-state values, $\bar{R} = \frac{\pi_{ss}}{\beta}$ is the steady-state nominal interest rate, and ε_t^R is a monetary policy shock. In addition, the fiscal authority finances a stochastic expenditure stream $g_y \bar{Y} \varepsilon_t^G$:

$$\ln(\varepsilon_t^G) = \rho_g \ln(\varepsilon_{t-1}^G) + u_t^G + \rho_{GA} u_t^A,$$

where we allow for a correlation between exogenous spending and innovations to total factor productivity.⁴⁹ For simplicity, we rule out government debt ($s_t = 0$), implying that all expenditures are financed by lump sum taxes; i.e. $g_y \bar{Y} \varepsilon_t^G = T_t$.

Competitive equilibrium: The competitive equilibrium is a path of 41 variables ($\Lambda, K, L, y, l, c, q, k, d, e, \pi, r_k, r_d, r_e, R, W, mc, o^{new}, o, \pi^*, Z_{p1}, Z_{p2}, Z_{p3}, Z_{w1}, Z_{w2}, Z_{w3}, \Delta_{p1}, \Delta_{p2}, \Delta_{p3}, \Delta_{p4}, \Delta_{w1}, \Delta_{w2}, \Delta_{w3}, \Delta_{p4}, W^*, i, \tilde{R}, \xi, \tilde{\xi}, \tilde{\psi}, \tau$) that satisfy the following 41 equations at each point in time, given initial conditions and the exogenous shock processes $\varepsilon^A, \varepsilon^B, \varepsilon^G, \varepsilon^I, \varepsilon^P, \varepsilon^R, \varepsilon^W, \varepsilon^\xi$.

⁴⁸Both ε_t^p and ε_t^w follow the standard shock process augmented by an moving average component, as in Smets and Wouters (2007).

⁴⁹This is a shortcut to take exports into account. Productivity innovations might raise exports in the data, and a way to capture it in a closed-economy model such as ours is to allow for $\rho_{GA} \neq 0$ as in Smets and Wouters (2007).

Household⁵⁰

$$\Lambda_t = \varepsilon_t^B (c_t - \iota c_{t-1})^{-\sigma_C} - \beta \iota E_t \left[\varepsilon_{t+1}^B (c_{t+1} - \iota c_t)^{-\sigma_C} \right] \quad (22)$$

$$E_t \left[\Lambda_{t+1} \frac{q_t r_{d,t+1} + (1 - q_t) \frac{\psi}{1-k_t} \pi_{t+1}}{\pi_{t+1}} \right] = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] \quad (23)$$

$$E_t \left[\Lambda_{t+1} \frac{q_t r_{e,t+1} - \xi_t \pi_{t+1}}{\pi_{t+1}} \right] = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] \quad (24)$$

$$\Lambda_t = \beta E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] \quad (25)$$

$$y_t = c_t + i_t + g_y \bar{Y} \varepsilon_t^G \quad (26)$$

Goods sector

$$\frac{L_t^d}{K_{t-1}} \frac{\alpha}{1 - \alpha} = \frac{r_{k,t}}{w_t} \quad (27)$$

$$mc_t = \frac{1}{A_t} \alpha^{-\alpha} r_{kt}^\alpha w_t^{1-\alpha} (1 - \alpha)^{\alpha-1} \quad (28)$$

$$\pi_t^* = \frac{\epsilon_p (1 + k_p)}{\epsilon_p (1 + k_p) - 1} \frac{Z_{p1,t}}{Z_{p2,t}} + \frac{k_p}{\epsilon_p - 1} (\pi_t^*)^{1+\epsilon_p(1+k_p)} \frac{Z_{p3,t}}{Z_{p2,t}} \quad (29)$$

$$Z_{p1,t} = (1 - \tau_{p,t}) \Lambda_t mc_t y_t \Delta_{p1,t}^{\epsilon_p(1+k_p)/(1-\epsilon_p(1+k_p))} + \beta \lambda_p E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{\epsilon_p(1+k_p)} Z_{p1,t+1} \right] \quad (30)$$

$$Z_{p2,t} = (1 - \tau_{p,t}) \Lambda_t y_t \Delta_{p1,t}^{\epsilon_p(1+k_p)/(1-\epsilon_p(1+k_p))} + \beta \lambda_p E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{\epsilon_p(1+k_p)-1} Z_{p2,t+1} \right] \quad (31)$$

$$Z_{p3,t} = \Lambda_t y_t + \beta \lambda_p E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{-1} Z_{p3,t+1} \right] \quad (32)$$

$$\Delta_{p1,t} = (1 - \lambda_p) (\pi_t^*)^{1-\epsilon_p(1+k_p)} + \lambda_p \Delta_{p1,t-1} \left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{\epsilon_p(1+k_p)-1} \quad (33)$$

$$1 = \frac{1}{1 + k_p} \Delta_{p1,t}^{1/(1-\epsilon_p(1+k_p))} + \frac{k_p}{1 + k_p} \Delta_{p2,t} \quad (34)$$

⁵⁰Note that R^d and R^e have been substituted out. The FOC w.r.t. labour is merged with the labour sector equations.

$$\Delta_{p2,t} = (1 - \lambda_p) \pi_t^* + \lambda_p \Delta_{p2,t-1} \left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{-1} \quad (35)$$

$$\Delta_{p3,t} = \frac{1}{1 + k_p} \Delta_{p1,t}^{\epsilon_p(1+k_p)/(1-\epsilon_p(1+k_p))} \Delta_{p4,t} + \frac{k_p}{1 + k_p} \quad (36)$$

$$\Delta_{p4,t} = (1 - \lambda_p) (\pi_t^*)^{-\epsilon_p(1+k_p)} + \lambda_p \Delta_{p4,t-1} \left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{\epsilon_p(1+k_p)} \quad (37)$$

$$A_t K_{t-1}^\alpha \left(\frac{L_t}{\Delta_{p3,t}} \right)^{1-\alpha} = \Delta_{p3,t} y_t \quad (38)$$

Labor sector

$$w_t^* = \frac{\epsilon_w (1 + k_w)}{\epsilon_w (1 + k_w) - 1} \frac{Z_{w1,t}}{Z_{w2,t}} + \frac{k_w}{\epsilon_w - 1} (w_t^*)^{1+\epsilon_p(1+k_p)} \frac{Z_{w3,t}}{Z_{w2,t}} \quad (39)$$

$$\begin{aligned} Z_{w1,t} &= \varepsilon_t^B \bar{L} L_t^{1+\sigma_L} w_t^{\epsilon_w(1+k_w)} (C_t - \iota C_{t-1})^{1-\sigma_c} \exp \left(\bar{L} \frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_L} \right) \Delta_{w1,t}^{\epsilon_w(1+k_w)/(1-\epsilon_w(1+k_w))} \\ &\quad + \beta \lambda_w E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_w} \bar{\pi}^{1-\gamma_w}} \right)^{\epsilon_w(1+k_w)} Z_{w1,t} \right] \end{aligned} \quad (40)$$

$$Z_{w2,t} = (1 - \tau_{w,t}) \Lambda_t L_t \left[w_t \Delta_{w1,t}^{1/(1-\epsilon_w(1+k_w))} \right]^{\epsilon_w(1+k_w)} + \beta \lambda_w E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_w} \bar{\pi}^{1-\gamma_w}} \right)^{\epsilon_w(1+k_w)-1} Z_{w2,t+1} \right] \quad (41)$$

$$Z_{w3,t} = (1 - \tau_{w,t}) \Lambda_t L_t + \beta \lambda_w E_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\gamma_p} \bar{\pi}^{1-\gamma_p}} \right)^{-1} Z_{w3,t+1} \right] \quad (42)$$

$$\Delta_{w1,t} = (1 - \lambda_w) \left(\frac{w_t^*}{w_t} \right)^{1-\epsilon_w(1+k_w)} + \lambda_w \Delta_{w1,t-1} \left(\frac{w_{t-1}}{w_t} \right)^{1-\epsilon_w(1+k_w)} \left(\frac{\pi_{t+1}}{\pi_t^{\gamma_w} \bar{\pi}^{1-\gamma_w}} \right)^{\epsilon_w(1+k_w)-1} \quad (43)$$

$$1 = \frac{1}{1 + k_w} \Delta_{w1,t}^{1/(1-\epsilon_w(1+k_w))} + \frac{k_w}{1 + k_w} \Delta_{w2,t} \quad (44)$$

$$\Delta_{w2,t} = (1 - \lambda_w) \left(\frac{w_t^*}{w_t} \right) + \lambda_w \Delta_{w2,t-1} \left(\frac{w_t}{w_{t-1}} \frac{\pi_{t+1}}{\pi_t^{\gamma_w} \bar{\pi}^{1-\gamma_w}} \right)^{-1} \quad (45)$$

$$\Delta_{w3,t} = \frac{1}{1 + k_w} \Delta_{w1,t}^{\epsilon_w(1+k_w)/(1-\epsilon_w(1+k_w))} \Delta_{w4,t} + \frac{k_w}{1 + k_w} \quad (46)$$

$$\Delta_{w4,t} = (1 - \lambda_w) \left(\frac{w_t^*}{w_t} \right)^{-\epsilon_w(1+k_w)} + \lambda_w \Delta_{w4,t-1} \left(\frac{w_t}{w_{t-1}} \frac{\pi_{t+1}}{\pi_t^{\gamma_w} \bar{\pi}^{1-\gamma_w}} \right)^{\epsilon_w(1+k_w)} \quad (47)$$

Government

$$R_t - \bar{R} = \phi_\pi \frac{\pi_{t+s}}{\bar{\pi}} + \phi_y \frac{y_{t+s}}{\bar{y}} + \phi_k \frac{k_{t+s}}{\bar{k}} + \rho (R_{t-1} - \bar{R}) \quad (48)$$

Capital producer

$$K_t = q_t \left(\omega_1 - \frac{\omega_2}{2} q_t \right) o_t + (1 - q_t) \theta \quad (49)$$

$$o_t = o_t^{new} + (1 - \delta) K_{t-1}$$

$$o_t^{new} = \varepsilon_T^I i_t \left(1 - \frac{\kappa}{2} \left(\frac{i_t}{i_{t-1}} - 1 \right)^2 \right) \quad (51)$$

$$Q_t \varepsilon_t^I \left[1 - S \left(\frac{i_t}{i_{t-1}} \right) - S' \left(\frac{i_t}{i_{t-1}} \right) \frac{i_t}{i_{t-1}} \right] - 1 = \beta E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \varepsilon_{t+1}^I Q_{t+1} S' \left(\frac{i_{t+1}}{i_t} \right) \left(\frac{i_{t+1}}{i_t} \right)^2 \right]. \quad (52)$$

Bank

$$q_t = 1 - \frac{\tilde{R}}{\tilde{\psi}_t} + \frac{\sqrt{\omega_2 (\tilde{R}_t - \tilde{\psi}_t) (\tilde{R}_t + 2\tilde{\xi}_t) (2\omega_1 \tilde{\psi}_t (\tilde{R}_t + \tilde{\xi}_t) + \omega_2 (\tilde{R}_t - \tilde{\psi}_t) (\tilde{R}_t + 2\tilde{\xi}_t))}}{\omega_2 \tilde{\psi}_t (\tilde{R}_t + 2\tilde{\xi}_t)} \quad (53)$$

$$k_t = \frac{\tilde{R}_t - \tilde{\psi}_t}{\tilde{R}_t + 2\tilde{\xi}_t} \quad (54)$$

$$\left(\omega_1 - \frac{\omega_2}{2} q_{t-1} \right) \frac{r_{k,t} + (1 - \delta) Q_t - \tau_t}{Q_{t-1}} - \frac{r_{d,t}}{\pi_{t+1}} (1 - k_t) - \frac{r_{e,t+1}}{\pi_{t+1}} k_t = 0 \quad (55)$$

$$\tau_t = \frac{Q_{t-1}^{\frac{1-q_{t-1}}{q_{t-1}}} \left(\psi - \theta \frac{r_{k,t} + (1-\delta)Q_t}{Q_{t-1}} \right)}{\omega_1 - \frac{\omega_2}{2} q_{t-1}} \quad (56)$$

$$\tilde{\xi}_t = \xi_t E_t [\Lambda_{t+1}] \quad (57)$$

$$\xi_t = \xi \varepsilon_t^\xi \quad (58)$$

$$\tilde{R}_t = E_t \left[\Lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] \quad (59)$$

$$\tilde{\psi}_t = \psi E_t [\Lambda_{t+1}] \quad (60)$$

$$o_t Q_t = e_t + d_t \quad (61)$$

$$k_t = e_t / (e_t + d_t) \quad (62)$$

Shock processes

$$\log(\varepsilon_t^B) = \rho_P \log(\varepsilon_{t-1}^B) + \sigma^B u_t^B \quad (63)$$

$$\log(\varepsilon_t^Q) = \rho_I \log(\varepsilon_{t-1}^Q) + \sigma^Q u_t^Q \quad (64)$$

$$\log(\varepsilon_t^\xi) = \rho_\xi \log(\varepsilon_{t-1}^\xi) + \sigma^\xi u_t^\xi \quad (65)$$

$$\log(\varepsilon_t^P) = (1 - \rho_P) \log(\bar{\varepsilon}_{t-1}^P) + \rho_P \log(\varepsilon_{t-1}^P) + \sigma^P (u_t^P + m_P u_{t-1}^P) \quad (66)$$

$$\log(\varepsilon_t^W) = (1 - \rho_W) \log(\bar{\varepsilon}_{t-1}^W) + \rho_W \log(\varepsilon_{t-1}^W) + \sigma^W (u_t^W + m_W u_{t-1}^W) \quad (67)$$

$$\log(\varepsilon_t^A) = \rho_A \log(\varepsilon_{t-1}^A) + \sigma^A u_t^A \quad (68)$$

$$\log(\varepsilon_t^R) = \rho_R \log(\varepsilon_{t-1}^R) + \sigma^R u_t^R \quad (69)$$

$$\log(\varepsilon_t^G) = \rho_G \log(\varepsilon_{t-1}^G) + \sigma^G u_t^G + \rho_{GA} \sigma^A u_t^A \quad (70)$$

Observational equations: The observation equations, linking the observed time series (left hand-side) to the variables in the non-linear model (right hand-side) are the following:

$$100\Delta \log\left(\frac{Y_t}{Y_{t-1}}\right) = 100\Delta \log\left(\frac{y_t}{y_{t-1}}\right) + 100\mu_y$$

$$100\Delta \log\left(\frac{C_t}{C_{t-1}}\right) = 100\Delta \log\left(\frac{c_t}{c_{t-1}}\right) + 100\mu_y$$

$$100\Delta \log\left(\frac{I_t}{I_{t-1}}\right) = 100\Delta \log\left(\frac{i_t}{i_{t-1}}\right) + 100\mu_y$$

$$100\Delta \log\left(\frac{W_t}{W_{t-1}}\right) = 100\Delta \log\left(\frac{w_t}{w_{t-1}}\right) + 100\mu_y$$

$$100\Delta \log\left(\frac{P_t}{P_{t-1}}\right) = 100\pi_t$$

$$100 \log \left(\frac{H_t}{\bar{H}} \right) = 100 \log \left(\frac{L_t}{\bar{L}} \right) + 100\mu_l$$

$$\left(\frac{R_t}{4} \right) = 100R$$

$$\tilde{E}_t = 100k_t$$

where \bar{H} are hours worked in 2009 and μ_l is a shift parameter. Since there is no growth in the model, we estimate the mean growth rate in the data μ_y . The equity ratio in the data \tilde{E}_t is transformed by taking deviations from its linear trend and adding back the mean.

Prior specifications: We fix parameters that are not identified to values commonly used in the literature. In particular, we choose a depreciation rate δ of 0.025, a steady-state wage markup $\bar{\varepsilon}^W$ of 1.05, a steady-state spending to GDP ratio g_y of 18%, a weight of labor in the utility function \bar{L} such that steady-state hours are equal to 1, and curvatures of the Kimball aggregator for goods and labor varieties of 10.

For all structural shocks, we employ a non-informative uniform distribution. The persistences of the shock processes are assumed to have a beta prior distribution centred at 0.5, and with standard deviation of 0.2. Following Smets and Wouters (2007), we further assume that the two markup shows have a moving average component.

The priors of the Taylor rule parameters are centred around very common values: the smoothing parameter has a Beta distribution with a mean of 0.75, while the responses to inflation and output are assumed to follow a Normal distribution with a mean of 1.5 and of $0.5/4 = 0.125$.

Since we use level data of the inflation rate and of the nominal interest rate, we choose the priors for the steady state of the inflation rate $\bar{\pi}$ and the real interest rate $1/\beta - 1$ to match the mean in the data, i.e. we assumed they follow a gamma distribution respectively centred around annualized values of 2.5% and 0.9.

The parameters affecting price and wage stickiness have a beta distribution centered at 0.5 with standard deviation of 0.1. Our prior is that prices and wages are reoptimized on average every 6 months, and that the degree of indexation to past inflation is only up to 50%. The steady-state price markup is assumed to be centred around 1.25, slightly above the steady-state wage markup.

We employ very common priors for all the parameters of the utility function. Habits are centred around 0.7, the intertemporal elasticity of substitution σ_c has a prior mean of 1.5, while the elasticity of labor supply σ_l has a prior mean of 2. The capital share in production has a prior mean of 0.3 while the investment adjustment costs parameter has a loose prior around 4.

For the discussion on the priors for the banking sector parameters, we refer to the main text.

Table 4: **Model estimation:** prior and posterior values

	parameter	prior shape	prior mean	prior std	post. mean	90% HPD interval	
	structural parameters						
μ_y	trend growth	norm	0.4	0.1	0.4264	0.3908	0.4618
μ_l	labor normalization	norm	0	2	-0.0938	-1.6569	1.4777
α	output share	norm	0.3	0.05	0.2001	0.1602	0.2395
$100\frac{1-\beta}{\beta}$	real rate in %	norm	0.25	0.1	0.427	0.2992	0.5485
$\bar{\varepsilon}^P$	price markup	norm	1.25	0.12	1.5068	1.3621	1.6523
$\bar{\pi}$	inflation in %	gamma	0.62	0.1	0.6263	0.4893	0.7616
ϕ_π	TR weight on inflation	norm	1.5	0.25	1.8723	1.5489	2.2003
ϕ_y	TR weight on output	norm	0.12	0.05	0.0198	-0.0348	0.0753
ρ	TR persistence	beta	0.75	0.1	0.8411	0.8057	0.8768
κ	investment adj. costs	norm	4	1.5	7.4584	5.5992	9.3376
ι	habits	norm	0.7	0.1	0.7774	0.7042	0.8532
σ_c	risk aversion	gamma	1.5	0.375	1.7362	1.2809	2.1939
σ_l	disutility from labor	gamma	2	0.75	2.0183	0.9726	3.0566
λ_p	price calvo parameter	beta	0.5	0.1	0.6206	0.5429	0.701
λ_w	wage calvo parameter	beta	0.5	0.1	0.8476	0.8099	0.8864
γ_p	price indexation	beta	0.5	0.15	0.1533	0.0537	0.2479
γ_w	wage indexation	beta	0.5	0.15	0.448	0.2066	0.6829
ξ	equity premium	norm	0.015	0.01	0.0213	0.0054	0.0348
θ	liquidation value	norm	0.5	0.1	0.7416	0.6425	0.8385
\bar{k}	equity ratio	norm	0.12	0.05	0.1231	0.1208	0.1254
	structural shock processes						
σ_A	stdev TFP	unif	0	10	0.3665	0.3172	0.414
σ_B	stdev preference	unif	0	10	3.4696	2.2271	4.6946
σ_G	stdev govt. spending	unif	0	10	2.2678	1.984	2.5382
σ_I	stdev investment	unif	0	10	4.7269	3.0495	6.3757
σ_P	stdev price markup	unif	0	1	0.1332	0.109	0.1574
σ_R	stdev monetary	unif	0	1	0.1164	0.1009	0.1315
σ_W	stdev wage markup	unif	0	10	0.4742	0.4088	0.5389
σ_ξ	stdev equity premium	unif	0	10	0.5805	0.199	1.0255
ρ_A	persistence TFP	beta	0.5	0.2	0.4623	0.3496	0.5765
ρ_B	persistence preference	beta	0.5	0.2	0.9004	0.8549	0.9486
ρ_G	persistence gov. spending	beta	0.5	0.2	0.9009	0.8471	0.9556
ρ_I	persistence investment	beta	0.5	0.2	0.1924	0.0357	0.3396
ρ_P	persistence price markup	beta	0.5	0.2	0.9772	0.9625	0.9925
ρ_R	persistence monetary	beta	0.5	0.2	0.9585	0.918	0.9967
ρ_W	persistence wage markup	beta	0.5	0.2	0.7721	0.6706	0.8734
ρ_ξ	persistence equity premium	beta	0.5	0.2	0.8156	0.7623	0.8699
$\rho_{G,A}$	correlation gov. spending & TFP	beta	0.5	0.2	0.6513	0.3835	0.9394
m_p	MA component of price markup	beta	0.5	0.2	0.7765	0.6826	0.875
m_w	MA component of wage markup	beta	0.5	0.2	0.9741	0.9516	0.9972

Appendix C: Proofs

The risk-taking channel for a generic expected return function

Consider the bank problem discussed in Section 2, but replace the expression for the expected return conditional on success $q_t (\omega_1 - \omega_2/2q_t)$ with the generic function $f(q_t)$.

Assume there exists an equilibrium $[\tilde{r}_{l,t}, q_t, k_t]$ under perfect competition that satisfies the following conditions: (1) the bank's choices are interior, i.e. $[k_t, q_t] \in [0, 1]^2$, (2a) the default probability is low relative to the parameters $\frac{q_t}{(1-q_t)}\tilde{\xi}_t \geq \tilde{R}_t - \tilde{\psi}_t$ or (2b) there is no deposit insurance $\tilde{\psi}_t = 0$, the conditional expected return function $f(q_t)$ satisfies (3) $f(q_t) \geq 0$, $f''(q_t) < 0$ and (4) $f'''(q_t) \leq 0$, $f''''(q_t) \leq 0$.

Note that assumption 2a), which is sufficient but by no means necessary and only needed for claim (e), is weak if we consider the empirically relevant section of the parameter space with a low equity premium (around 0.0x), a real rate just above 1 (1.0x) and high level of deposit insurance (0.x) and high repayment probabilities (0.9x). Assumption 3 is straightforward as it guarantees a meaningful risk return trade-off with an interior solution. Assumption 4 is another sufficient but non-necessary condition.

We prove that if such a solution exists, then:

- (a) risk taking is excessive: $q_t < \arg\max f(q_t)$,
- (b) the safety of assets q_t is a positive function of \tilde{R}_t : $\frac{\partial q_t}{\partial \tilde{R}_t} > 0$,
- (c) the equity ratio k_t is a positive function of \tilde{R}_t : $\frac{\partial k_t}{\partial \tilde{R}_t} > 0$,
- (d) the expected return on investment is a positive function of \tilde{R}_t :

$$\frac{\partial f(q_t(\tilde{R}_t)) + (1-q_t(\tilde{R}_t))\theta}{\partial \tilde{R}_t} > 0,$$
- (e) the expected return on investment is a concave function of \tilde{R}_t :

$$\frac{\partial^2 f(q_t(\tilde{R}_t)) + (1-q_t(\tilde{R}_t))\theta}{\partial \tilde{R}_t^2} < 0.$$

For a generic return function $f(q_t)$ the bank's objective function at the second stage is:

$$\max_{q_t} f(q_t) \tilde{r}_{l,t} - q_t \tilde{r}_{d,t} (1 - k_t)$$

Deriving this problem with respect to q_t yields the following first-order condition, which by concavity is necessary and sufficient:

$$f'(q_t) \tilde{r}_{l,t} = \tilde{r}_{d,t} (1 - k_t) \quad (71)$$

Notice that this condition implies $f'(q_t) > 0$ ($k_t \in (0, 1]$ by assumption, $\tilde{r}_{d,t} > 0$ by the deposit supply schedule, and $\tilde{r}_{l,t} > 0$ by the zero-profit condition). Notice further that in a frictionless world, e.g. without limited liability, the bank's risk choice would satisfy $q_t^{opt} = \arg\max f(q_t) + (1 - q_t)\theta$, i.e. $f'(q_t^{opt}) = \theta$. Since we have assumed above that

the recovery value is smaller than the deposit insurance cap, which in turn is smaller than the cost of deposits by lemma 1, we have: $\tilde{r}_{l,t}\theta < \tilde{\psi}_t < \tilde{r}_{d,t}(1 - k_t)$. Combining this with equation (71) and the frictionless optimality condition and rearranging, we obtain $f'(q_t) > f'(q_t^{opt})$. Given $f''(q_t) < 0$ this implies excessive risk taking, i.e. $q_t < q_t^{opt}$ (claim(a)).

Since the deposit supply schedule must hold in equilibrium, we can rewrite this condition as:

$$f'(q_t)\tilde{r}_{l,t} - \frac{\tilde{R}_t(1-k_t) + (1-q_t)\tilde{\psi}_t}{q_t} = 0 \quad (72)$$

Equation (72) implicitly defines $\hat{q}_t(k_t)$. Using the implicit function theorem we find that, $f'(q_t^{opt}) = \theta < \tilde{r}_{d,t}(1 - k_t) = f'(q_t)\tilde{r}_{l,t}$

$$\frac{\partial q_t}{\partial k_t} = \frac{-q_t \tilde{R}_t}{(1-k_t)\tilde{R}_t - \tilde{\psi}_t + q_t^2 \tilde{r}_{l,t} f''(q_t)}$$

At the first stage the maximization problem is:

$$\max_{k_t} f(q_t)\tilde{r}_{l,t} - q_t\tilde{r}_{d,t}(1 - k_t) - q_t k_t \tilde{r}_{e,t} .$$

which, using the deposit and equity supply schedules $\tilde{r}_{d,t} = \frac{\tilde{R}_t - \frac{1-q_t}{1-k_t}\tilde{\psi}_t}{q_t}$ $\tilde{r}_{e,t} = \frac{\tilde{R}_t + \tilde{\xi}_t}{q_t}$, can be written as:

$$\max_{k_t} f(\hat{q}_t)\tilde{r}_{l,t} + (1 - q_t)\tilde{\psi}_t - k_t\tilde{\xi}_t - \tilde{R}_t .$$

The corresponding FOC is:

$$\left(f'(\hat{q}_t)\tilde{r}_{l,t} - \tilde{\psi}_t\right)\frac{\partial q_t}{\partial k_t} - \tilde{\xi}_t . \quad (73)$$

Finally, the zero-profit condition can in expectations be written as:

$$f(\hat{q}_t)\tilde{r}_{l,t} + (1 - q_t)\tilde{\psi}_t - k_t\tilde{\xi}_t - \tilde{R}_t . \quad (74)$$

Equations (72), (73), (74) implicitly define q_t , k_t and $\tilde{r}_{l,t}$. Solving the latter two equations for k_t and $\tilde{r}_{l,t}$ we obtain:

$$k_t = \frac{(-\tilde{\xi}_t\tilde{R}_t + q_t\tilde{R}_t\tilde{\psi}_t + \tilde{\xi}_t\tilde{\psi}_t)f(q_t) - q_t(\tilde{R}_t - (1-q_t)\tilde{\psi}_t)(\tilde{R}_t f'(q_t) + q_t\tilde{\xi}_t f''(q_t))}{-\tilde{\xi}_t(\tilde{R}_t f(q_t) - q_t(\tilde{R}_t f'(q_t) + q_t\tilde{\xi}_t f''(q_t)))} \quad (75)$$

$$\tilde{r}_{l,t} = \frac{(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t)}{\tilde{R}_t f(q_t) - q_t(\tilde{R}_t f'(q_t) + q_t\tilde{\xi}_t f''(q_t))} \quad (76)$$

Plugging these equations into (72) and rearranging, we obtain the following equation, which implicitly defines q_t :

$$(\tilde{R}_t + \tilde{\xi}_t) \frac{\tilde{R}_t \tilde{\psi}_t f(q_t) - (\tilde{R}_t(\tilde{R}_t + \tilde{\xi}_t) - ((1-q_t)\tilde{R}_t + \tilde{\xi}_t)\tilde{\psi}_t) f'(q_t) - q_t \tilde{\xi}_t (\tilde{R}_t - (1-q_t)\tilde{\psi}_t) f''(q_t)}{-\tilde{\xi}_t (\tilde{R}_t f(q_t) - q_t (\tilde{R}_t f'(q_t) + q_t \tilde{\xi}_t f''(q_t)))} = 0$$

We can simplify this condition further by multiplying by the denominator and dividing by $(\tilde{R}_t + \tilde{\xi}_t) \tilde{R}_t$:

$$F(q_t, R_t) \equiv \frac{\tilde{R}_t \tilde{\psi}_t f(q_t) - (\tilde{R}_t(\tilde{R}_t + \tilde{\xi}_t) - ((1-q_t)\tilde{R}_t + \tilde{\xi}_t)\tilde{\psi}_t) f'(q_t) - q_t \tilde{\xi}_t (\tilde{R}_t - (1-q_t)\tilde{\psi}_t) f''(q_t)}{\tilde{R}_t} = 0 \quad (77)$$

Using the implicit function theorem on equation (77), we find that :

$$\frac{\partial q_t}{\partial R_t} = - \frac{\frac{\partial F}{\partial R_t}}{\frac{\partial F}{\partial q_t}} \quad (78)$$

where

$$\begin{aligned} \frac{\partial F}{\partial R_t} &= \frac{(\tilde{R}_t^2 + \tilde{\xi}_t \tilde{\psi}_t) f'(q_t) + (1-q_t) q_t \tilde{\xi}_t \tilde{\psi}_t f''(q_t)}{-\tilde{R}_t^2} \\ \frac{\partial F}{\partial q_t} &= \frac{(\tilde{R}_t - (1-q_t)\tilde{\psi}_t) ((\tilde{R}_t + 2\tilde{\xi}_t) f''(q_t) + q_t \tilde{\xi}_t f'''(q_t))}{-\tilde{R}_t} \end{aligned}$$

Using our assumptions on f , the parameters and assuming an interior solution, it is obvious that $\frac{\partial F}{\partial q_t} > 0$. How about the $\frac{\partial F}{\partial R_t}$?

To get the sign of $\frac{\partial F}{\partial R_t}$, we solve 77 for $f(q_t)$:

$$f(q_t) = \frac{(\tilde{R}_t(\tilde{R}_t + \tilde{\xi}_t) - ((1-q_t)\tilde{R}_t + \tilde{\xi}_t)\tilde{\psi}_t) f'(q_t) + q_t \tilde{\xi}_t (\tilde{R}_t - (1-q_t)\tilde{\psi}_t) f''(q_t)}{\tilde{R}_t \tilde{\psi}_t}$$

and plug this expression into the equations (75) and (76) for k_t and $\tilde{r}_{l,t}$:

$$k_t = \frac{f'(q_t)(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t) + f''(q_t) q_t \tilde{\xi}_t (\tilde{R}_t - (1-q_t)\tilde{\psi}_t)}{\tilde{R}_t ((\tilde{R}_t + \tilde{\xi}_t) f'(q_t) + q_t \tilde{\xi}_t f''(q_t))^2} \quad (79)$$

$$\tilde{r}_{l,t} = \frac{(\tilde{R}_t + \tilde{\xi}_t) \tilde{\psi}_t}{(\tilde{R}_t + \tilde{\xi}_t) f'(q_t) + q_t \tilde{\xi}_t f''(q_t)} \quad (80)$$

Since in equilibrium $\tilde{r}_{l,t} > 0$ and since the numerator of $\tilde{r}_{l,t}$ is positive, it must hold that its denominator is also positive:

$$(\tilde{R}_t + \tilde{\xi}_t) f'(q_t) + q_t \tilde{\xi}_t f''(q_t) > 0 \quad (81)$$

Similarly, since $k_t > 0$ and since the denominator of k_t is positive, the numerator must be positive too:

$$f'(q_t)(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t) + f''(q_t) q_t \tilde{\xi}_t (\tilde{R}_t - (1-q_t)\tilde{\psi}_t) > 0 \quad (82)$$

Since $f' > 0$ and $f'' < 0$ we can conclude from the previous inequality that for any

$[x_1, x_2] \in \mathbb{R}^2$ it must hold that $f'(q_t)x_1 + f''(q_t)x_2 > 0$ if

$$\frac{x_1}{x_2} \geq \frac{(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t)}{q_t \tilde{\xi}_t (\tilde{R}_t - (1 - q_t)\tilde{\psi}_t)}. \quad (83)$$

We now test this condition for the numerator of $\frac{\partial F}{\partial \tilde{R}_t}$:

$$\frac{\tilde{R}_t^2 + \tilde{\xi}_t \tilde{\psi}_t}{(1 - q_t)q_t \tilde{\xi}_t \tilde{\psi}_t} \leq \frac{(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t)}{q_t \tilde{\xi}_t (\tilde{R}_t - (1 - q_t)\tilde{\psi}_t)}.$$

Rearranging, multiplying only by positive values, yields:

$$0 \leq -\tilde{R}_t(\tilde{R}_t - (1 - q_t)\tilde{\psi}_t) - q_t \tilde{\psi}_t \tilde{\xi}_t - (\tilde{R}_t(1 - q_t)\tilde{\psi}_t + (1 - q_t)\tilde{\psi}_t \tilde{\xi}_t) \frac{q_t \tilde{\psi}_t}{(\tilde{R}_t - (1 - q_t)\tilde{\psi}_t)}$$

The RHS is obviously negative since from the proposition that every deposit insurance cap will be exceeded, it follows that $\tilde{R}_t > \tilde{\psi}_t$. Hence the condition $\frac{\tilde{R}_t^2 + \tilde{\xi}_t \tilde{\psi}_t}{(1 - q_t)q_t \tilde{\xi}_t \tilde{\psi}_t} \geq \frac{\tilde{R}_t + \tilde{\xi}_t}{q_t \tilde{\xi}_t}$ is satisfied and we can conclude that the numerator of $\frac{\partial F}{\partial \tilde{R}_t}$ is positive. Hence $\frac{\partial F}{\partial \tilde{R}_t} < 0$ and therefore $\frac{\partial q_t}{\partial \tilde{R}_t} > 0$ (claim (b)).

Equation (79) defines $k_t = \mathfrak{K}(q_t, \tilde{R}_t)$. Its derivative is given by

$$\frac{\partial k_t}{\partial \tilde{R}_t} = \frac{\partial \mathfrak{K}}{\partial \tilde{R}_t} + \frac{\partial \mathfrak{K}}{\partial q_t} \frac{\partial q_t}{\partial \tilde{R}_t}$$

where

$$\begin{aligned} \frac{\partial \mathfrak{K}}{\partial \tilde{R}_t} &= \frac{((f'(q_t))^2 \tilde{R}_t^2 + 2f'(q_t)(f'(q_t) + f''(q_t)(1 - q_t)q_t)\tilde{R}_t \tilde{\xi}_t + (f'(q_t) + f''(q_t)q_t)(f'(q_t) - f''(q_t)(1 - q_t)q_t)\tilde{\xi}_t^2)\tilde{\psi}_t}{\tilde{R}_t^2 (f''(q_t)\tilde{\xi}_t + f'(\tilde{R}_t + \tilde{\xi}_t))^2} \\ \frac{\partial \mathfrak{K}}{\partial q_t} &= \frac{q_t \tilde{\xi}_t f''(q_t) \tilde{\psi}_t (2(\tilde{R}_t + \tilde{\xi}_t)f'(q_t) + q_t \tilde{\xi}_t f''(q_t))}{\tilde{R}_t (f''(q_t)\tilde{\xi}_t + f'(\tilde{R}_t + \tilde{\xi}_t))^2} \end{aligned}$$

From (83) it is immediately obvious that the numerator of $\frac{\partial \mathfrak{K}}{\partial q_t}$ is negative, hence $\frac{\partial \mathfrak{K}}{\partial q_t} < 0$. After division by $\tilde{\psi}_t$, the numerator of $\frac{\partial \mathfrak{K}}{\partial \tilde{R}_t}$ can be rewritten as:

$$((\tilde{R}_t + \tilde{\xi}_t)f'(q_t) + q_t \tilde{\xi}_t f''(q_t))^2 - (f''(q_t))^2 q_t^3 \tilde{\xi}_t^2 - f''(q_t)f'(q_t)q_t^2 \tilde{\xi}_t (2\tilde{R}_t + \tilde{\xi}_t)$$

Since the first term is positive and larger than the absolute value of the second term we can see that $\frac{\partial \mathfrak{K}}{\partial \tilde{R}_t} > 0$. Hence we have shown that $\frac{\partial k_t}{\partial \tilde{R}_t} > 0$ (claim (c)).

Applying the implicit function theorem a second time on equation (77), we can find the following expression for the second derivative of q_t :

$$\frac{\partial^2 q_t}{\partial \tilde{R}_t^2} = \frac{\left(\frac{\partial^2 F}{\partial \tilde{R}_t \partial q_t} + \frac{\partial^2 F}{\partial q_t^2} \frac{\partial q_t}{\partial \tilde{R}_t}\right) \frac{\partial F}{\partial \tilde{R}_t} - \left(\frac{\partial^2 F}{\partial \tilde{R}_t \partial q_t} \frac{\partial q_t}{\partial \tilde{R}_t} + \frac{\partial^2 F}{\partial \tilde{R}_t^2}\right) \frac{\partial F}{\partial q_t}}{\left(\frac{\partial F}{\partial q_t}\right)^2} \quad (84)$$

where:

$$\frac{\partial^2 F}{\partial \tilde{R}_t \partial q_t} = \frac{(\tilde{R}_t^2 + 2(1 - q_t)\tilde{\xi}_t\tilde{\psi}_t) f''(q_t) + q_t(1 - q_t)\tilde{\xi}_t\tilde{\psi}_t f'''(q_t)}{-\tilde{R}_t}$$

$$\frac{\partial^2 F}{\partial q_t^2} = \frac{\tilde{\psi}_t (q_t\tilde{\xi}_t f'''(q_t) + (\tilde{R}_t + 2\tilde{\xi}_t) f''(q_t)) + (f'''(q_t)q_t\tilde{\xi}_t + f'''(q_t)(\tilde{R}_t + 3\tilde{\xi}_t)) (\tilde{R}_t - (1 - q_t)\tilde{\psi}_t)}{-\tilde{R}_t}$$

$$\frac{\partial^2 F}{\partial R_t^2} = \frac{2(f'(q_t) + f''(q_t)(1 - q)q)\tilde{\xi}_t\tilde{\psi}_t}{\tilde{R}_t^3}$$

since $f'' < 0$ and $f''' \leq 0$ $f'''' \leq 0$ and all parameters are positive, it is obvious that $\frac{\partial^2 F}{\partial R_t \partial q_t} > 0$ and $\frac{\partial^2 F}{\partial q_t^2} > 0$. The term $\frac{\partial^2 F}{\partial R_t^2}$ is less straightforward. A sufficient condition for $\frac{\partial^2 F}{\partial R_t^2} > 0$ can be found using condition (83):

$$\frac{1}{(1 - q_t)q_t} \geq \frac{(\tilde{R}_t + \tilde{\xi}_t)(\tilde{R}_t - \tilde{\psi}_t)}{q_t\tilde{\xi}_t(\tilde{R}_t - (1 - q_t)\tilde{\psi}_t)}$$

which simplifies to:

$$\frac{q_t}{(1 - q_t)}\tilde{\xi}_t \geq \tilde{R}_t - \tilde{\psi}_t$$

Given the signs of the terms in (84) we have finally verified that:

$$\frac{\partial^2 q_t}{\partial \tilde{R}_t^2} = \frac{((+) + (+)(+))(-) - ((+)(+) + (+))(+)}{(+) } < 0$$

Under alternative assumption (2b) the expression for $\frac{\partial^2 q_t}{\partial \tilde{R}_t^2}$ simplifies to:

$$\frac{\partial^2 q_t}{\partial \tilde{R}_t^2} = - \frac{f'(q_t)(-2f''(q_t)f'''(q_t)q_t\tilde{\xi}_t - 2(f''(q_t))^2(\tilde{R}_t + 2\tilde{\xi}_t) + f'(q_t)(f'''(q_t)q_t\tilde{\xi}_t + f'''(q_t)(\tilde{R}_t + 3\tilde{\xi}_t)))}{(f'''(q_t)\tilde{\xi}_t + f''(q_t)(\tilde{R}_t + 2\tilde{\xi}_t))^3}$$

which is negative without further conditions.

Using the signs of the derivatives of q_t and the fact that $f'(q_t) > f'(q_t^{opt}) = \theta$, we can finally determine the slope and curvature of the expected return on the bank's investment.

$$\begin{aligned} \frac{\partial[f(q_t) + (1 - q_t)\theta]}{\partial \tilde{R}_t} &= \underbrace{(f'(q_t) - \theta)}_{+} \underbrace{\frac{\partial q_t}{\partial \tilde{R}_t}}_{+} > 0 \\ \frac{\partial^2[f(q_t) + (1 - q_t)\theta]}{\partial \tilde{R}_t^2} &= \underbrace{(f'(q_t) - \theta)}_{+} \underbrace{\frac{\partial^2 q_t}{\partial \tilde{R}_t^2}}_{+} + \underbrace{f''(q_t)}_{-} \underbrace{\frac{\partial q_t}{\partial \tilde{R}_t}}_{+} < 0 \end{aligned}$$

This completes the proof of claims (d) and (e).

Notice that the quadratic functional form we assumed for $f(q_t)$ in the model section satisfies assumptions (3) and (4) and we focused on interior solutions (assumption (1)).

Therefore claims (1), (2), (3) and (4) in propositions 1 and 2 hold. Furthermore, claim (5) in proposition 1 holds since assumption (2a) is satisfied. Finally, to see that claim (5) in proposition 2.1 holds independent of assumption (2a) and (2b), consider the solution for q_t

$$q_t = 1 - \frac{\tilde{R}}{\tilde{\psi}_t} + \frac{\sqrt{\omega_2(\tilde{R}_t - \tilde{\psi}_t)(\tilde{R}_t + 2\tilde{\xi}_t)(2\omega_1\tilde{\psi}_t(\tilde{R}_t + \tilde{\xi}_t) + \omega_2(\tilde{R}_t - \tilde{\psi}_t)(\tilde{R}_t + 2\tilde{\xi}_t))}}{\omega_2\tilde{\psi}_t(\tilde{R}_t + 2\tilde{\xi}_t)}$$

The second derivative of this expression is:

$$\frac{\partial^2 q}{\partial \tilde{R}_t^2} = - \frac{\omega_1\omega_2 \left\{ 2\omega_2 (\tilde{R}_t - \tilde{\psi}_t)^3 \tilde{\xi}_t (\tilde{R}_t + 2\tilde{\xi}_t) + \omega_1\tilde{\psi}_t \dots \right.}{(\tilde{R}_t + 2\tilde{\xi}_t) \left\{ \omega_2(\tilde{R}_t - \tilde{\psi}_t)(\tilde{R}_t + 2\tilde{\xi}_t) [2\omega_1\tilde{\psi}_t(\tilde{R}_t + \tilde{\xi}_t) + \omega_2(\tilde{R}_t - \tilde{\psi}_t)(\tilde{R}_t + 2\tilde{\xi}_t)] \right\}^{2/3}} \left. \dots \left[\tilde{R}_t^4 + 2\tilde{R}_t (4\tilde{R}_t^2 - 3\tilde{R}_t\tilde{\psi}_t + 2\tilde{\psi}_t^2\tilde{\xi}_t) + (12\tilde{R}_t - 4\tilde{R}_t\tilde{\psi}_t + 5\tilde{\psi}_t^2) \tilde{\xi}_t^2 + 8\tilde{R}_t + 4\tilde{\psi}_t\tilde{\xi}_t^3 + 16\tilde{\xi}_t^4 \right] \right\}$$

Both the numerator and denominator are positive, so $\frac{\partial^2 q}{\partial \tilde{R}_t^2} < 0$. Hence $\frac{\partial^2 [f(q_t) + (1 - (q_t))\theta]}{\partial \tilde{R}_t^2} < 0$. ■

Deposits in excess of insurance

The proof is by contradiction: Assume that there exists an equilibrium with no excess profits where the bank would issue so little deposits that the promised repayment $r_{d,t}$ would be lower than the cap on deposit insurance $\psi/(1 - k_t)\pi_{t+1}$.⁵¹ In this case the deposit rate $r_{d,t}$ would be equal to the risk-free rate R_t .

The second-stage maximization problem of the bank would then be:

$$\max_{q_t \in [0,1]} f(q_t)\tilde{r}_{l,t} - q_t\tilde{R}_t(1 - k_t)$$

and its solution \hat{q}_t is implied by $f'(q_t)\tilde{r}_{l,t} = \tilde{R}_t(1 - k_t)$. The first-stage maximization problem is:

$$\max_{k_t \in [0,1]} V(k_t) = f(\hat{q}_t)\tilde{r}_{l,t} - \hat{q}_t\tilde{R}_t(1 - k_t) - (\tilde{\xi}_t + \tilde{R}_t)k_t$$

\hat{q}_t can either be a corner or an interior solution. If \hat{q}_t is a corner solution, the first-stage objective function of the bank is obviously decreasing in k_t , hence $k_t = 0$ is optimal. If \hat{q}_t is an interior solution, the first derivative of the first-stage objective function is:

$$-\tilde{\xi}_t - \tilde{R}_t(1 - \hat{q}_t)$$

Since $\hat{q}_t \in [0, 1]$, this derivative is negative for all $k_t \in [0, 1]$, i.e. the objective function is again decreasing in k . Hence the solution to the first-stage problem is $k_t = 0$. Optimality

⁵¹For simplicity, we abstract from the possibility that the cap is binding for some states of the future but not for others, which would be possible due to the inconsistency between the timing of inflation and the nominal deposit rate. Note that this distinction disappears under certainty equivalence or first-order approximation.

with full insurance therefore requires the bank to use only deposits. This contradicts our initial assumption. This result implies that any insurance cap smaller than 100% would be exceeded by the deposit liabilities in case of default. Depositors are therefore never fully insured.

Notice that for a cap to be effective in the sense of ruling out full insurance equilibria, the cap has to be low enough. Formally speaking, it needs to hold that $r_{d,t}(1 - k_t) > \tilde{\psi}_t$ even under full insurance, i.e. $\tilde{R}_t > \tilde{\psi}_t$. ■

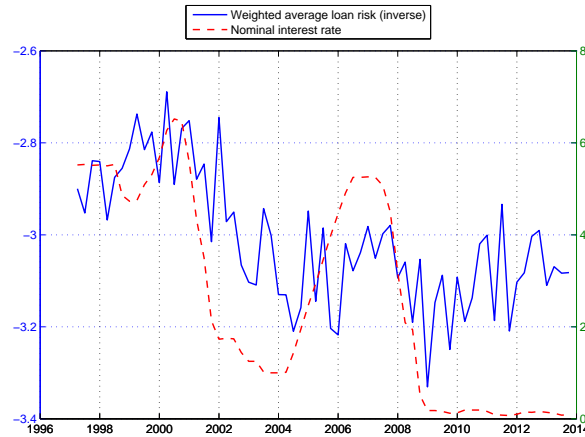
Appendix D: Empirical motivation - the asset risk-taking channel in the US

To motivate our theoretical analysis we provide here additional empirical evidence on the existence of the asset risk-taking channel in the US. Our starting point is a classical small-scale VAR that includes inflation, output, a measure of bank risk taking and the effective federal funds rate, taken as the monetary policy instrument. Output is measured by real GDP growth, while inflation is defined as the log change in the GDP deflator. The transformation of the variables matches that used for the DSGE estimation, and is described in appendix A. Identification of the monetary policy shock is achieved through sign restrictions that do not involve the measure of bank risk, and are in this sense agnostic. We find that an unexpected decrease in the risk-free interest rate causes a persistent increase in bank risk taking, a result robust to using a recursive identification scheme.

Measuring bank risk-taking behavior There are many notions of asset risk. One can distinguish between ex-ante, ex-post and realized asset risk. The former is the risk perceived by the bank when making a loan or buying an asset. Banks can influence this class of risk directly, when making their investment decisions (the ex-ante risk choice). On the other hand, the ex-post risk of a bank's balance sheet is also affected by unforeseen changes in asset riskiness, that take place after origination and are largely outside the banks' influence. Lastly, the payoff ultimately paid by an asset is a materialization of the former two types of uncertainty (realized asset risk). In this paper we focus on active risk taking, that is the level of ex-ante risk that intermediaries choose. Ex-ante bank risk taking is however largely unobservable. Inferring it from realized risk (e.g. loan losses) is hardly possible with aggregate data. Inference from the spread between some measure of bank funding costs and loan rates neglects the fact that this spread not only reflects default risk but also incorporates a liquidity premium and the markup, which are likely to be affected by the same variables that influence the risk choice. Instead, we use a survey-based proxy for bank risk-taking, which is provided by the US Terms of Business Lending Survey and consists in the internal risk rating assigned by banks to newly issued loans. In this survey, available from 1997-Q1 onwards, 400 banks report the

volume of loans originated in first week of the mid month each quarter, grouped by the internal risk rating. This rating varies between 1 and 5, with 5 being the maximum level of risk. Following Dell’Ariccia et al. (2014), we construct a weighted average loan risk series, using as weights the value of loans in each risk category.⁵² We plot this measure of risk, together with the nominal risk-free interest rate, in figure 3. We can see that low levels of the monetary policy instrument tend to be associated with higher level of bank risk (lower or decreasing levels in the blue line). This is particularly evident in the period between 2000 and 2006, when the monetary policy stance in the US was deemed to be particularly lax.⁵³

Figure 3: Bank risk taking and nominal interest rate: The risk measure (solid blue line, left axis) is redefined such that a decrease is associated with higher risk-taking of the banking sector, matching the definition in the theoretical model discussed later. The nominal interest rate (dashed line, right axis) is the effective federal funds rate.



Identification strategy and results We estimate a VAR on US data from 1997:Q2, the start of the survey-based proxy for risk taking, to 2007:Q3, as described in appendix A⁵⁴. The lag length is chosen to be 1, as indicated by the BIC information criterion.

We identify an unexpected monetary policy shock by using a set of sign restrictions that are robust across a variety of general-equilibrium models. In particular, we assume that an expansionary monetary policy shock decreases the nominal risk-free interest rate, and increases inflation and output, both at the time of the shock and in the quarter immediately after. Note that the response of inflation ensures that this shock is identified separately from a productivity shock, as summarized in table 5.

⁵²The average loan risk is a perfect measure for bank risk taking if we assume that the volume of loans is constant. Else, banks could also minimize their risk exposure by reducing the quantity of loans as their average quality goes down. While the correlation between risk and loan volume growth is slightly negative, it is not significant at a 10% significance level. For a more in-depth discussion of the data we refer to Buch et al. (2014).

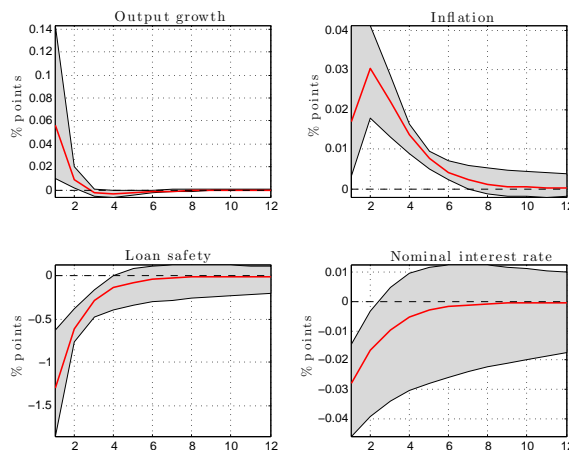
⁵³See, for instance, the discussion in Taylor (2007).

⁵⁴We have decided to cut the zero-lower bound period, but our results still hold when the latest available data are used.

Table 5: **Sign restriction identification scheme:** restrictions are assumed to hold on impact and on the following period.

SHOCK	variables		
	y	π	R
MONETARY POLICY SHOCK	+	+	−
TOTAL FACTOR PRODUCTIVITY SHOCK	+	−	−

Figure 4: **Monetary policy shock on bank risk-taking:** Impulse responses over a 3-year horizon, identified through the sign restriction scheme in Table 1. Error bands correspond to 90% confidence intervals reflecting rotation uncertainty. Loan safety is defined as the inverse of the average loan risk rating, standardized to take values between 0 and 100. The remaining variables are annualized. See text for further details.

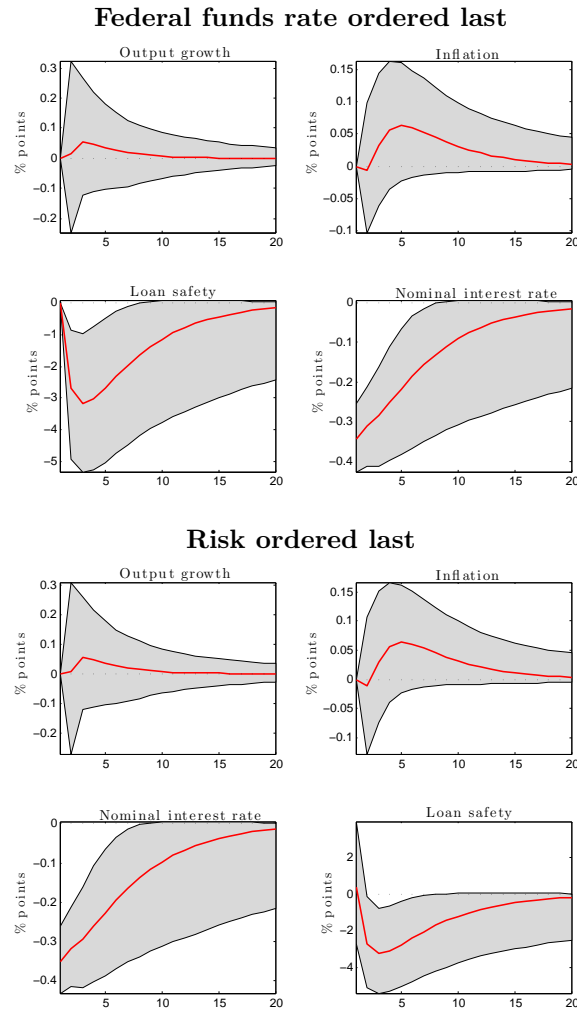


The response of bank asset risk to an expansionary monetary policy shock is shown in figure 4. An unexpected decrease in the monetary policy interest rate is followed by a moderate macroeconomic expansion: output growth increases for less than a year, while inflation displays a longer reaction of about two years. The results are compatible with the existence of a risk-taking channel in the US: a fall in the nominal interest rate leads in fact to a decrease in the ex-ante proxy for the safety of banks' assets, i.e. banks issue riskier loans. Interestingly, the implied responses of the nominal interest rate and the risk measure are approximately proportional.

We assess the robustness of our results by employing a recursive identification scheme as in Angeloni et al. (2015). Output and inflation are always ordered before the nominal interest rate, implicitly assuming that they do not react to interest rate shocks, nor to shocks coming from the banking sector. More controversial is the ordering of the nominal interest rate and of the loan risk rating. We experiment with both orderings and show in figure 5 that results do not depend on the ordering chosen and confirm the results obtained using sign restrictions.⁵⁵

⁵⁵Our measure of risk taking does not take into account loan quantities. For example, riskier lending may go hand in hand with lower loan volumes which may leave the expected loss volume unchanged.

Figure 5: **An expansionary monetary policy shock - Recursive identification scheme.** Error bands correspond to 90% confidence intervals obtained by bootstrap. Loan safety is defined as the inverse of the average loan risk rating, standardized to take values between 0 and 100. The remaining variables are annualized. See text for further details.



When adding loan volumes to the VAR we find that the response of volumes is not significantly different from zero with the mean showing no clear pattern in either direction, while the other IRFs remain unchanged.

Chapter II

Austerity to save the banks? A quantitative model of sovereign default with endogenous default costs and a financial sector

1 Introduction

In the recent European debt crisis many governments faced a dilemma: Enforce austerity in times of crisis to generate the surplus needed to repay their debt or default on their debt obligations and face the financial disruptions entailed. In many cases it seemed hard to believe that austerity policies were the less painful choice and some countries eventually defaulted, most notably Greece. Yet the Greek experience, just like the recent defaults by Russia or Argentina, confirmed that default does indeed entail a significant cost: In all three cases the financial sector suffered major losses, which lead to severe repercussions in the real economy.

These dramatic events brought an old question to the forefront of the policy and academic debate: Why should sovereign governments repay their debt, even though there is no institution that can force them to? This paper proposes a quantitative model of opportunistic sovereign default with endogenous default costs reflecting the recent experiences in Greece and elsewhere to answer this question.

In the model debt is issued by the government of a small open economy. This economy is populated by households, banks and firms, which produce output from labor and capital. The banks issue deposits to the household and invest into loans to firms and sovereign bonds. Their capacity to intermediate private savings within the economy is constrained by a financial friction, which limits their leverage. Besides the domestic banks, foreign investors are also active in the sovereign debt market. This debt is commitment free, that is the government, which maximizes the domestic households' utility, can default on its debt in a nondiscriminatory fashion whenever it finds it opportune to do so. Since this model does not rely on any exogenous costs of default, not repaying the foreign held debt per-se is a free lunch. Yet since defaulting on domestic banks generates a financial crisis in the domestic economy, default becomes costly to the benevolent government. Besides

I would like to thank Arpad Abraham, Luigi Boccola, Nils Gornemann, Enrique Mendoza, Evi Pappa, Cesar Sosa-Padilla and Leopold von Thadden as well as my colleagues at European University Institute and the ECB and the organizers and participants of the 2014 ZICE workshop for useful comments. I thank the ECB for funding my research stay in Frankfurt.

the output losses due to the financial crisis, default furthermore is followed by a period during which the government prefers not to issue new debt given the current prices. In other words, the government self excludes itself from international capital markets. The model is hence capable of simultaneously explaining both the output costs of default as well as the breakdown of international lending after default. These costs constitute a commitment device, which explains why governments mostly repay and hence why they *can* borrow. Furthermore, the existence of capital in combination with the frictions in domestic intermediation drive a wedge between the actual return on capital and the return perceived by the household. This gives rise to an arbitrage motive, which explains why governments *want* to borrow.

The model is calibrated to Greek data. Besides inheriting the quantitative features of the RBC production structure, the model is successful at replicating both the empirical magnitude of the output losses of default as well as estimates of the distribution of the duration of the international capital market shutdown. Furthermore, I use the model to argue that a recently debated policy proposal to reduce the exposure of European banks to domestic sovereign debt may have severe unintended consequences: As the banks exposure is reduced banks do not only become more resilient to sovereign debt crisis. Such a policy furthermore reduces the degree to which the government can commit repayment of its debt. Given a 20% reduction in the banks' exposure, the model predicts a reduction in the total amount of foreign debt sustainable by about one half. If implemented abruptly, this reform would trigger default with a high probability.

This theory is motivated by a large body of empirical evidence. De Paoli et al. (2006), Sturzenegger and Zettelmeyer (2007), Acharya et al. (2011), Bolton and Jeanne (2011), Reinhart and Rogoff (2011b), Sosa-Padilla (2012) and Gennaioli et al. (2014) all document the link between sovereign default and financial crisis. Gennaioli et al. (2014) for example compile an international panel comprising data on the macro economy, public debt and financial institutions' balance sheets, covering 110 default episodes. They document that domestic banks hold significant amounts of domestic sovereign debt and that defaults often precede financial crises. Furthermore, they provide evidence that credit extended by banks to the private sector drops significantly in case of default and that this drop is stronger, the more exposed banks are to sovereign debt. Consequentially, they find that higher shares of domestic debt reduce the probability of default.

This paper is related to a recent field of theoretical literature that analyzes sovereign default in quantitative models, based on the Eaton and Gersovitz (1981) framework. Motivated by empirical observations this literature initially simply *assumed* that default on external sovereign debt is costly, due to capital market exclusion and output costs (see Arellano (2008) or Aguiar and Gopinath (2006)). Subsequently, researchers attempted to endogenously *explain* the costs of default. For example, D'Erasmo and Mendoza (2012) show that the distributional effects of default can explain default and repayment, while

Guembel and Sussman (2009) outline a theory of political support in favor of or against default. In Mendoza and Yue (2012) default leads to market exclusion not only for the government but also for firms, which need the international markets to finance inter-period working capital. This paper adds to this literature by proposing a quantitative model where the costs of default arise from the exposure of the financial system to government debt.

In doing so it provides 3 novel conceptual insights, all of which are found to be quantitatively relevant: First, the introduction of physical capital not only bridges the gap between the sovereign default literature and the RBC model⁵⁶, it furthermore yields a new theory for why governments borrow. Virtually all the external sovereign default literature to date relies on the assumption that the borrowing country is impatient relative to the world interest rate to explain why the country borrows. I show that capital as a domestic savings technology together with the financial friction implies a permanent difference between the return on domestic capital and the world interest rate, creating an opportunity for arbitrage. Second, this is the first model on sovereign default to generate market exclusion as an equilibrium outcome. The existing literature typically simply introduces exclusion by assumption, often referring to it as a coordinated punishment strategy by lenders⁵⁷. In many cases where the output costs of default are modeled endogenously, they actually result from this assumption (e.g. Mendoza and Yue (2012)). On the contrary, in the model I propose market exclusion arises as an optimal choice: During the financial crisis the marginal return on domestic investment is depressed and falls below the world interest rate. Hence, the gains from trade vanish temporarily, borrowing becomes relatively expensive and abstention from external borrowing is optimal. Third, this paper provides a new quantitative theory of the costs of sovereign default, where sovereign default leads to a breakdown of intermediation resulting in a period of underinvestment.

Four recent working papers also endogenize output costs by introducing a financial sector, many of which have been concurrent work. Like in this paper, in these models the bank is exposed to sovereign debt. Sosa-Padilla (2012) develops a very stylized model in which banks' wealth directly serves as an input to production, without providing intermediation services. In Engler and Grosse Steffen (2016), Mallucci (2014) and Perez (2015) and local banks trade on inter-banking markets, and sovereign default brings these markets to collapse: in the first and second case because sovereign bonds are necessary as collateral to facilitate interbank lending and this collateral disappears in case of default, in the last because the interbank-leverage is constrained. The collapse of the interbank markets then leads to efficiency losses. While similar in spirit, a number of features in my model are unique. First, all these models have in common that the bank needs to finance

⁵⁶Joo (2014) and Gornemann (2015) also introduce capital into a default model.

⁵⁷Kletzer (1994) has criticized this argument, since it would be profitable for each individual lender to deviate from the punishment strategy.

the labor bill in advance through intra-period loans. In my model on the contrary, the bank finances capital. This gives rise to a more realistic bank balance sheet and, as explained above, generates a novel motive for borrowing. Second, the model I propose gives rise to exclusion endogenously as a consequence of the financial crisis. On the contrary, all of these papers simply assume exclusion and, apart from Sosa-Padilla (2012), rely on this assumption to generate the financial crisis and the corresponding output costs of default.

The idea that sovereign default inflicts costly damage to the financial sector is also developed in a more stylized manner in 3 periods models like Acharya et al. (2011), Basu (2010), Bolton and Jeanne (2011), Brutti (2011), Gennaioli et al. (2014), Erce (2012) and Mayer (2011). Due to their 1 shot nature, none of these models can speak about market exclusion or be tested quantitatively.

At the same time this paper is also related to the macro literature on the role of frictions in the financial sector, for which the handbook chapter of Gertler and Kiyotaki (2010) is representative. Yet, while the shocks to the banks' balance sheets in these model result from the banks' lending to the real economy, I focus on the shocks resulting from their lending to the government. Boccia (2016) analyzes sovereign default and its consequences in the context of the former model, yet he considers the default decision as exogenous.

The rest of the paper is structured as follows: The next section lays out a quantitative model of sovereign default with a 3 sector production economy and discusses the key assumptions. Section 3 explains the mechanisms of the model, before the computational strategy and the calibration are discussed in section 4. Section 5 presents the main quantitative results and section 6 uses the model to evaluate the effects of a policy that reduces the exposure of banks to domestic sovereign debt. The last section concludes.

2 The model

This chapter proposes a dynamic small open economy model of sovereign default, where the only costs of default stem from the exposure of the domestic banking sector to sovereign risk. The domestic private economy is made up of 3 sectors, namely households, banks and firms and is governed by a benevolent government. The domestic economy is open and can borrow from international lenders (the rest of the world).

2.1 The household

There is a continuum of mass 1 of identical households, which each make a consumption-labor-savings choice. The household is risk averse and can invest into bank deposits d_h and loans to firms k_h . Bank deposits are promises to pay 1 unit of the final good tomorrow, which can be bought at price r . Loans are promises to repay R' units of the final good tomorrow, which can be bought at price $1 + \xi$, where ξ is an exogenous variable, which

reflects transaction costs and which follows a stochastic process discussed later.⁵⁸ Both assets are safe and loans can't be shorted.⁵⁹ The household's disposable income is the sum of the value of his assets carried over from the last period $d_h + Rk_h$, his labor income Wl_h and the lump sum dividend and profit payments from the banks Div and the firms π , minus a lump sum tax T . The household has rational expectations and chooses labor, consumption and investment to maximize his lifetime utility, taking prices and aggregate states as given.

The household's problem hence is:

$$V_h(d, \Omega) = \max_{c, d'_h, k'_h, l_h} u(c, l) + \beta E_{|\Omega} V_h(d', \Omega')$$

$$st.$$

$$c + rd'_h + (1 + \xi)k'_h + T = d_h + Rk_h + Wl_h + Div + \pi \quad (1)$$

$$k'_h \geq 0$$

where Ω denotes the aggregate state vector. For instantaneous utility I choose a conventional CARA function with additively separable disutility from labor, as for example in Boccia (2016):

$$u(c, l) = \frac{c^{1-\gamma}}{1-\gamma} + \chi \frac{l^{1+\nu}}{1+\nu}$$

The corresponding first order conditions are necessary and sufficient:

$$u_c(c, l) r = \beta E_{|\Omega} [u_c(c', l')] \quad (2)$$

$$u_c(c, l)(1 + \xi) = \beta R' E_{|\Omega} [u_c(c', l')] + \lambda_h \quad (3)$$

$$0 = \min \{ \lambda_h, k'_h \} \quad (4)$$

$$u_c(c, l)W = u_l(c, l) \quad (5)$$

2.2 The banks

The bank is the key agent in this economy. Banks intermediate the savings from the households to the firm more efficiently than if the household invests directly, yet they are constraint by a financial friction that gives rise to a leverage constraint. It is this constraint that ultimately makes default costly for the government.

There is continuum of mass 1 of identical banks, each of which is run by a banker. (In this subsection, I drop the index b to simplify notation). The banker is a small member of the household, hence he uses the household's stochastic discount factor to value the

⁵⁸As will become clear later, the fact that banks do not face these transaction costs makes intermediated lending desirable.

⁵⁹Explicitly state contingent contracts or default are ruled out by assumption.

payoff of his activities.

The banker enters each period with a portfolio of 1 period assets and liabilities (his balance sheet) chosen last period. In particular, he holds (1) k units of firm loans which pay the ex-ante fixed gross return R ; (2) b units of government bonds which return 1 unit if repaid ($Rep = 1$), otherwise they return nothing ($Rep = 0$); and (3) he owes depositors d . The net cash flow of these 3 assets defines the bank's pre-dividend equity e .

$$e = Rk + bRep - d$$

From this net cash-flow, the banker chooses how much to take home to his household, or, more formally speaking, how much dividend div to pay to his shareholders. The rest he carries into the period on the balance sheet as post dividend equity $e - div$. Given the post-dividend equity, the banker next chooses his optimal portfolio of assets and liabilities: he can invest in loans to firms k' at price 1, buy government bonds b' at price q , and sell deposits d' at price r . This yields the following end-of-period balance sheet:

Assets	Liabilities
Loans k'	Deposits $r \cdot d'$
Government bonds $q \cdot b'$	Post dividend equity $e - div$

which can equivalently be expressed by the balance sheet equation:

$$qb' + k' = e - div + rd'$$

Note that, from the lenders point of view, loans by banks are different than loans by the household - think of the latter as corporate bonds or entrepreneurial self-financing. While the bank pays 1 unit to provide 1 unit of financing to the firm, the household pays $1 + \xi$ to provide 1 unit of financing to the firm. This cost difference stems from the superior screening and monitoring technology that the bank possesses. Bank intermediated financing is hence more efficient, which provides an explanation for why banks exist. As mentioned before ξ is allowed to vary over the cycle. In particular I assume that ξ is related to total factor productivity (TFP) ω according to $\xi = \bar{\xi}\omega^{-\hat{\xi}}$.

Besides the balance sheet equation, the banker's choice is constrained by several other constraints. The first constraint is crucial. Following Iacoviello (2005) I assume that bankers face a commitment problem: Between periods, i.e. after having invested into loans and bonds, but before the revelation of next period TFP, bankers can abandon the bank. In that case the depositors would take over the bank, but they would only be able to recover a fraction θ of the cash-flow, since they lack the special skills to run the bank. This option gives the banker the opportunity to renegotiate the deposit contract. Assuming that the banker has all bargaining power, the banker makes depositors a take-it-or-leave-it offer: Either depositors agree to reducing the deposit repayment to a level

that leaves them indifferent between accepting or rejecting the offer, or the banker will abandon the bank. Of course depositors anticipate this, which is why, *ex ante*, they will not accept any contract that is not renegotiation-proof. This translates into a leverage constraint, requiring the bank to hold a certain minimum of expected pre-dividend equity tomorrow (weighted by the depositors' expected marginal utility):

$$E_{|\Omega} [u_c(c')] d' \leq E_{|\Omega} [u_c(c') \theta (R e p' b' + R' k')]$$

Second, the banker's dividend choice is constrained below, it can not be smaller than a certain $(1 - \eta)$ fraction of equity.

$$div \geq (1 - \eta) e$$

This assumption, which is common in the macro literature on financial frictions⁶⁰, makes sure that the leverage constraint introduced above bites: One on hand, the banker can't simply raise new equity should he require more. On the other hand, the fact that he constantly has to payout a positive dividend ensures that he also can't overcome the equity constraint over time by accumulating retained earnings. At the same time it is not completely at odds with reality: Banks seem to have difficulties raising capital especially in times of crisis and they generally try to smooth their dividends.⁶¹

Third, the bank has to hold a fixed fraction of government bonds on the balance sheet:

$$\frac{b' q}{b' q + k'} = \psi$$

Fourth, next period equity must not be negative.

$$R k \geq d$$

Finally, there is no short selling of assets.

$$b' \geq 0, k' \geq 0$$

The bank has rational expectations and maximizes the discounted value of dividends, taking the discount factor, the prices and the law of motion of the aggregate state as given. Her objective function reads

$$V_b(d, b, k, \Omega) = \max_{b', k', d', div} div + E_{|\Omega} \left(\beta \frac{u_c(c')}{u_c(c)} \right) V_b(d', b', k', \Omega')$$

After substituting out *div*, rearranging and abstracting from the short selling con-

⁶⁰E.g. Gertler and Kiyotaki (2010), who motivate the same assumption by retiring bankers.

⁶¹The constraint implies that the dividend per (book) equity ratio (also known as dividends per share) is bounded below by a fraction of the *gross* return on (book) equity. Since the *gross* return is fairly stable across the cycle, even if the *net* return varies strongly, dividends per share are fairly stable.

straints, which cannot be binding in any equilibrium with banking, and the non-negativity constraint⁶², the banks problem is:

$$\begin{aligned}
V_b(d, b, k, \Omega) &= \max_{b', k', d'} (e + rd' - qb' + k') u_c(c) + \beta E_{|\Omega} [V_b(d', b', k', \Omega')] \\
&\quad st. \\
E_{|\Omega} [u_c(c')d'] &\leq E_{|\Omega} [u_c(c')\theta(Rep' b' + R' k')] \\
(1 - \eta)e &\leq e + rd' - qb' - k' \\
b'q &= \psi(b'q + k') \\
\text{where } e &= Rk + bRep - d
\end{aligned}$$

The solution of the bank's dynamic problem can conveniently be characterized as I show next.

Proposition 1: The vector of choice variables $[b', d', k'] \in (\mathbb{R}_0^+)^3$ is a solution to the banks problem if and only if together with the vector of multipliers $[\lambda_1, \lambda_2, \lambda_3] \in (\mathbb{R}_0^+)^3$ it solves the system of first order conditions

$$\partial \mathcal{L} / \partial d' : 0 = ru_c(c) - \beta E_{|\Omega} [V_{be}(e', \Omega')] - \lambda_1 E_{|\Omega} [u_c(c')] + \lambda_2 r \quad (6)$$

$$\partial \mathcal{L} / \partial b' : 0 = -qu_c(c) + \beta E_{|\Omega} [Rep' V_{be}(e', \Omega')] + \lambda_1 E_{|\Omega} [Rep' u_c(c')] \theta - \lambda_2 q + q\lambda_3(1 - \psi) \quad (7)$$

$$\partial \mathcal{L} / \partial k' : 0 = -u_c(c) + \beta RE_{|\Omega} [V_{be}(e', \Omega')] + \lambda_1 R' E_{|\Omega} [u_c(c')] \theta - \lambda_2 - \lambda_3 \psi \quad (8)$$

$$\partial \mathcal{L} / \partial \lambda_1 : 0 = \min \left\{ E_{|\Omega} [u_c(c)\theta(Rep' b' + R' k')] - E_{|\Omega} [u_c(c')d'] , \lambda_1 \right\} \quad (9)$$

$$\partial \mathcal{L} / \partial \lambda_2 : 0 = \min \{ \eta e - qb' - k' + rd' , \lambda_2 \} \quad (10)$$

$$\partial \mathcal{L} / \partial \lambda_3 : 0 = b'q - (b'q + k) \psi \quad (11)$$

where $V_{be}(e', \Omega')$, the derivative of the value function wrt. equity, is given by $V_{be}(e', \Omega') = u_c(c') + \eta \lambda_2'$.

Proof: Five steps are necessary to arrive to this result. First, note that the value function of the bank $V_b(d, b, k, \Omega)$ can be summarized as a function $V_b(e, \Omega)$ of only one endogenous state variable $e = Rk + bRep - d$ since (b, d, k) only enter the problem as this linear combination. Second, the solution of the problem of the bank $[b', d', k']$ and its value function $V_b(e, \Omega)$ is linear in e . To proof this second claim, assume that $V_b(e', \Omega')$ is linear in e' . Be x a solution to the bank's problem given e and Ω . Denote the associated value function by $\tilde{V}_b(e, \Omega)$. Then we can conclude that αx is optimal given αe for any $\alpha \in \mathbb{R}^+$ since both the objective and the constraints are linear in both e and x . (If there existed $\alpha x'$ that is feasible given αe such that $\tilde{V}_b(\alpha x', \Omega) > \tilde{V}_b(\alpha x, \Omega)$ then by linearity x'

⁶²The model will be parametrized such that this abstraction remains irrelevant.

would also be feasible given e and by homogeneity it would dominate x .) Therefore the solution of the banks problem - given a linear $V_b(e', \Omega')$ - is linear in e . Furthermore, by linearity the value of this solution $\tilde{V}_b(e, \Omega)$ also needs to be linear in e . Since the same reasoning applies to the value function of all the following periods, the initial assumption that $V_b(e', \Omega')$ is linear in e' must hold. Third, the first and second result together imply $V_b(e', \Omega')$ is linear in $[b', d', k']$. Fourth, given the third result it is obvious that both the constraints and the objective of the optimization problem are affine functions. Hence, the first order conditions are necessary and sufficient. Fifth, to determine the derivative of the value function w.r.t equity $V_{be}(e', \Omega')$ we can apply the envelope theorem for maximization problems with inequality constraints: $V_{be}(e', \Omega') = u_c(c') + \eta\lambda'_2$. ■

The second part of the proof furthermore implies that the distribution of equity among banks does not matter, which allows us to represent the banking sector by a representative bank. The last step furthermore implies that the marginal value of equity $V_{be}(e', \Omega')(e', \Omega')$ does not depend on e' . This allows us to eliminate all references to the value function from the FOCs and to characterize the bank's optimal choice in terms of policy functions alone, which is convenient for the computations.

2.3 The firms

There is a continuum of mass 1 of firms. The dynamic problem of the firm can be broken down to a 2 period problem. In the first period the firm borrows k_f units from the banks and households at a fixed interest rate R' , which she transforms into capital for the next period. In the second period the firm hires l_f units of labor from the households and produces. The firm uses a standard Cobb-Douglas production function with capital share α and stochastic TFP ω and capital depreciates at rate δ . TFP ω follows a log-normal AR(1) process with persistence ρ and variance σ^2

$$\log(\omega) = \rho \log(\omega_{-1}) + \sigma \varepsilon$$

Like the bank the firm manager is part of the household, to whom any profits or losses π are rebated lump sum and whose stochastic discount factor the firm manager uses. His sequential optimization problem is

$$\max_{k'_f} E_{|\Omega} \left[\left(\beta \frac{u_c(c')}{u_c(c)} \right) \left(\omega' (k'_f)^\alpha (\hat{l}(\omega'))^{1-\alpha} + (1 - \delta)k'_f - R'k'_f - W(\omega')\hat{l}(\omega')' \right) \right]$$

s.t

$$\hat{l}(\omega') = \arg\max_{l'_f} \omega' (k'_f)^\alpha (l'_f)^{1-\alpha} + (1 - \delta)k'_f - R'k'_f - W(\omega')l'_f$$

The first order conditions determining the loan rate and the wage are necessary and sufficient:

$$R' = E_{|\Omega} \left[u_c(c') \left(\omega' \alpha k_f^{\alpha-1} l_f^{1-\alpha} + (1 - \delta) \right) \right] / E_{|\Omega} [u_c(c')] \quad (12)$$

$$W = \omega(1 - \alpha) k_f^\alpha l_f^{1-\alpha} \quad (13)$$

2.4 The foreign investors

The model describes a small open economy. The rest of the world is represented by perfectly competitive risk neutral deep pocket foreign investors, who demand any asset inelastically that pays their expected rate of return $1/\bar{q}$ (i.e. the world interest rate). I assume that lending to domestic private agents requires local know-how and is not possible for the foreign investors, but they can invest in the 1 period government bond. Foreign demand for government bonds, denoted by B_x , is hence given by

$$B_x \in \begin{cases} 0 & \text{if } E_{|\Omega} R e p' / q < 1/\bar{q} \\ [0, \infty) & \text{if } E_{|\Omega} R e p' / q = 1/\bar{q} \end{cases} \quad (14)$$

2.5 Market clearing

This completes the discussion of the private agents' problems. We are ready to close the private economy by the corresponding market clearing conditions. From now I will denote aggregate choice variables by capital letters, i.e. $\int_0^1 x_i di = X_I$ where i is the respective agent's index and I its type. Labor market clearing requires

$$L_F = L_H \quad (15)$$

Loan and deposit market clearing implies

$$K'_F = K'_H + K'_B \quad (16)$$

$$D'_B = D'_H \quad (17)$$

Finally, bond markets clear when the total government debt issuance \bar{B}' equals the total demand by foreigners and local banks

$$B'_B + B'_X = \bar{B}' \quad (18)$$

and then by Walras' law the goods markets clear as well.

This concludes the set up of the private economy. Before we move to the problem of the

government, which will choose its tax, borrowing and repayment policy, it is convenient to define the notion of a private equilibrium, given a certain set of government policies $T(\Omega)$, $\bar{B}(\Omega)$, $Rep(\Omega)$.

Definition 1: Private equilibrium Given certain state dependent government policy functions $T(\Omega)$, $\bar{B}'(\Omega)$, $Rep(\Omega)$ a *stationary private equilibrium* consists of a set of state dependent policy functions $C(\Omega)$, $L_H(\Omega)$, $D'_H(\Omega)$, $K'_H(\Omega)$, $D'_B(\Omega)$, $B'_B(\Omega)$, $K'_B(\Omega)$, $K'_F(\Omega)$, $L_F(\Omega)$, $B'_X(\Omega)$ price functions $r(\Omega)$, $q(\Omega)$, $R'(\Omega)$, $W(\Omega)$ and shadow price functions $\lambda_1(\Omega)$, $\lambda_2(\Omega)$, $\lambda_3(\Omega)$, $\lambda_H(\Omega)$ such that the maximization problems of the household, the bank and the firm are solved, the foreign lenders' demand is satisfied, markets clear and such that the policy functions imply the state transition function that underlies the agents' expectations.

Call the vector of policy, price and shadow price functions $X(\Omega)$. Denote the functional equation system over the domain Ω , which is defined by equations (2) to (18) after replacing all individual variables with the corresponding aggregate variables⁶³, by $\mathcal{F}(X, \Omega) = 0$. Then, since all first order conditions are necessary and sufficient, an equivalent way of defining a *stationary private equilibrium* is: $X(\Omega)$ such that $\mathcal{F}(X, \Omega) = 0$ given any $T(\Omega)$, $\bar{B}(\Omega)$, $Rep(\Omega)$.

What are the state variables in this problem? First, we need to keep track of the exogenous state variable ω . Second, we need to account for the endogenous choices of the representative household and the bank, which have an inter-temporal dimension. This encompasses the loans extended by the household K_H ; the deposits issued by the bank and bought by the household, D ; the loans extended by the bank K_B ; the bonds bought by the bank B_B and the loan rate set ex ante R . Furthermore Ω encompasses the endogenous variable B_X , the amount of outstanding sovereign debt held abroad, because this variable influences the government policy choices, that are discussed next. The state vector is hence $\Omega = [\omega, K_H, D, B_B, K_B, R, B_X]$.

2.6 The government

The model economy is governed by a benevolent government, which chooses its actions such as to maximize the households' utility. It needs to finance some fixed government expenditures G and it can do so by taxing the household through the lump sum tax T ⁶⁴ and issuing 1 period government bonds \bar{B}' , which are a promise to repay 1 unit tomorrow and are traded at price q .⁶⁵ These bonds are sold on anonymous markets, where both

⁶³This change reflects the equilibrium requirement that individual and aggregate decisions coincide. In particular the bank's, firm's and the household's choice variables need to be replaced by their aggregate counterparts, i.e. capital letters.

⁶⁴Following the literature, I assume lump sum taxation. This simplifies the model by avoiding a motive for distortion smoothing.

⁶⁵Explicitly state contingent contracts are ruled out.

local banks and international investors can buy them. While it is known who buys the bonds (banks buy B'_B , foreign lenders invest B'_F), the government cannot discriminate buyers, neither at the time of issuance nor at the time of repayment. Sovereign debt is non-enforceable and the government can not commit; the government is free to choose to repay its debt ($Rep = 1$) or to fully default on its obligations ($Rep = 0$) at the beginning of each period after observing the realization of the TFP shock. In case of default there is no direct punishment: in particular foreign lenders do not refrain from lending, and no direct output costs arise. Yet I assume that following default the government effectively can not save abroad, because its foreign assets would else be seized by the creditors.⁶⁶ This gives rise to the following government budget constraint:

$$q \cdot \bar{B}' + T \geq B \cdot Rep + G \quad (19)$$

$$B_F \geq 0$$

It is important to note that the government is the only agent who is 'big' and hence fully understands the impact of each agents decision, including its own, on the equilibrium. This means it understands how its own decisions influence the choices of households, banks, firms and foreign lenders. While no direct default costs are assumed, the government understands that defaulting erodes the banks' equity, reduces their capacity to intermediate and hence leads to lower capital and therefore less production and income for the household. This is how default becomes costly to the government - despite the absence of any direct punishment - and hence why it may choose to repay for certain regions in the state space ex post. Furthermore, the government understands that its decision how much new debt to issue influences the expectations of everyone on the probability of default tomorrow. That means the government takes into account the effect of its actions on all private variables, including the bond price.

We can summarize the government's problem as follows in the notation common in the default literature:

$$V_G(\Omega) = \max_{Rep \in \{0,1\}} \{Rep V_{GR} + (1 - Rep) V_{GD}\}$$

⁶⁶In the optimization problem of the government below, I enforce this constraint also before default. This simplifies matters and is innocuous since in equilibrium the constraint only binds after default for the part of the parameter space that we are interested in.

where the value of repayment is

$$\begin{aligned}
V_{GR}(\Omega) &= \max_{T, \bar{B}', X} V_H(\Omega | Rep = 1) \\
&st. \\
q \cdot \bar{B}' + T &\geq \bar{B} + G \\
B_F &\geq 0 \\
0 &= \mathcal{F}(X, \Omega)
\end{aligned}$$

and the value of default is

$$\begin{aligned}
V_{GD}(\Omega) &= \max_{T, \bar{B}', X} V_H(\Omega | Rep = 0) \\
&st. \\
q \cdot \bar{B}' + T &\geq G \\
B_F &\geq 0 \\
0 &= \mathcal{F}(X, \Omega)
\end{aligned}$$

Note that the condition $0 = \mathcal{F}(X, \Omega)$ constrains the governments choices to be consistent with a private equilibrium.

Since there is no direct punishment for default, the value functions $V_{GR}(\Omega)$ and $V_{GD}(\Omega)$ are equivalent, and we can write the problem more concisely:

$$V_G(\Omega) = \max_{Rep, T, \bar{B}', X} V_H(\Omega) \quad (20)$$

st.

$$q \cdot \bar{B}' + T \geq \bar{B} \cdot Rep + G \quad (21)$$

$$B_F \geq 0 \quad (22)$$

$$0 = \mathcal{F}(X, \Omega) \quad (23)$$

Definition 2: Full equilibrium: A *stationary equilibrium* in this economy is defined by a set of value functions $V^H(\Omega)$, $V^G(\Omega)$, government policy functions $T(\Omega)$, $\bar{B}'(\Omega)$, $Rep(\Omega)$, private policy functions $C(\Omega)$, $L_H(\Omega)$, $D'_H(\Omega)$, $K'_H(\Omega)$, $D'_B(\Omega)$, $B'_B(\Omega)$, $K'_B(\Omega)$, $K'_F(\Omega)$, $L_F(\Omega)$, $B'_X(\Omega)$, price functions $r(\Omega)$, $q(\Omega)$, $R'(\Omega)$, $W(\Omega)$ and shadow price functions $\lambda_1(\Omega)$, $\lambda_2(\Omega)$, $\lambda_3(\Omega)$ such that the governments problem (20) is solved subject to its budget constraint (21), the non-negativity constraint (22) and the equilibrium conditions of the private economy (23). Appendix A summarizes the government problem.

Intuitively, one can think about the equilibrium as an infinitely repeated game. Each period the government moves first choosing the new debt and tax levels and whether

to default, anticipating correctly the reactions of the private agents, which in turn have correct expectations about the future. Yet another way to think about the full equilibrium is a private equilibrium that is associated with a set of time-consistent (i.e. commitment free) optimal government policy functions.

2.7 Discussion of key assumptions

No selective default and no bank bail outs The assumptions that the government can not choose to default selectively and can not bail out domestic banks after default are key to generate the endogenous costs of default that sustain an equilibrium with external debt. If the government could choose to default only on foreign lenders, or equivalently to inject capital selectively into domestic banks after defaulting, external default would be costless and hence always ex-post optimal. Since foreigners would anticipate that, no foreign lending could be sustained. Theoretically, both assumptions can be justified by secondary markets, as Broner et al. (2010) show. In anonymous markets, if the government planned to default selectively (or to bail some lenders out) the defaulted upon would simply sell to the exempted (bailed out). The non-selectivity of default is plausible also empirically. While there have been a few cases of selective default, they were never with respect to the *holder* but to the *currency* (Moody's (2008)) or the *legislation* (Reinhart and Rogoff (2011a)) of the bond. It seems rather hard to target the *holders* of the bond well. The empirical case for no bailouts seems a bit harder to make, given that we do see that government defaults are often accompanied by bank bailouts. I nevertheless follow the literature (e.g. Gertler and Kiyotaki (2010), Sosa-Padilla (2012), Engler and Grosse Steffen (2016)) and rule them out. This must be understood as an approximation for the fact that these rescue packages (1) typically do not fully compensate for the full default losses, (2) are politically costly⁶⁷, and (3) are subject to complicated legal constraints.

Full default In line with most of the quantitative sovereign default literature in general and the endogenous default cost literature more specifically, I assume that the government can only choose between full repayment or full default, even though this assumption is counterfactual - most defaults are partial (see e.g. Sturzenegger and Zettelmeyer (2008)). Furthermore, in my model the optimal haircut ex post would generally not be 100%, because the costs of default depend on the amount that banks lose, relative to their current equity. However, giving the country full discretion over the haircut seems unrealistic as well. A realistic representation of the default process would be to assume that a government can choose to either repay fully or enter debt renegotiations. At the

⁶⁷Political costs can arise because the population perceives that banks were “accomplices” in bringing upon the dire situation the country finds itself in or because they are provided by an external agent like the IMF who enforces some conditionality.

end of these, a proportional haircut is applied the country's debt and the remainder repaid. Yet this haircut is not simply a choice of the defaulting country, but is the outcome of a lengthy and ex-ante highly uncertain renegotiation process.⁶⁸ Given that debt renegotiation is not the focus of this paper, it is therefore reasonable to approximate this haircut as exogenous and potentially stochastic. In order to maintain comparability with virtually all the existing sovereign default literature, I choose to set the haircut to 100%. However, the mechanism of the model would not be altered if one would assume a smaller and/or stochastic haircut.

Bond holding requirement Another important assumption is that banks have to hold a certain fraction of their assets in domestic bonds. This assumption is crucial for the model, since banks would generally not find it optimal to invest into bonds that pay the world interest rate, which is generally lower than the return on domestic capital. At the same time their exposure is necessary to sustain any foreign lending. This simple constraint is intended to capture in a reduced form are all sorts of forces that induce banks to invest heavily into domestic sovereign bonds, such as financial regulation, financial repression⁶⁹ or liquidity considerations⁷⁰. Whatever the exact nature of these forces, its is a well documented empirical fact that banks exhibit a strong home bias and are heavily exposed to domestic sovereign debt: Sosa-Padilla (2012) and Gennaioli et al. (2014) report that on average 22% or 12% respectively of the financial sector's assets in each country in their samples are domestic sovereign debt.

International lending only to the government I assume that only the government can trade financial contracts with the rest of the world. Neither households, nor banks, nor firms can invest in or borrow from the rest of the world. This is a simplifying assumption, which could generally be relaxed as long as foreign lenders have some disadvantage in direct investment. Yet it is at the same time not an implausible approximation for many less financially integrated economies, where private borrowing makes up for only a small part of the total net foreign asset position, such as Greece for example.

Riskless loans The fact that I model loans as riskless and have the firm absorb the aggregate risk may appear unusual. Many papers with banking sectors like Gertler and Kiyotaki (2010) assume that instead bank equity buffers these shock. First note that in the absence of a friction at the level of the bank, both assumptions are equivalent. Given there are frictions they are not. In the real world most of banks assets are loans, that have a flat repayment profile as long as they are not defaulted upon. I abstract from firm

⁶⁸See Yue (2010a) for an extension of the Arellano model explicitly modeling the renegotiation process.

⁶⁹Notice that in particular regulation and repression could be considered policy variables as well. I abstract from these considerations. For a model with optimal financial repression see Chari et al. (2015).

⁷⁰For a model with a liquidity motive for holding government bonds see Engler and Grosse Steffen (2013).

default both for simplicity and because I want to isolate the effects of government default. Assuming that banks carry some of the risk associated with firm investment would change the quantitative properties of the model, since it would introduce a negative correlation between the innovation in TFP and bank equity and would hence reduce the degree to which default incentives increase in TFP.

Direct lending by the household In the model direct lending from the household to the firm - think of corporate bonds but also entrepreneurial self finance - is generally less efficient due to the costs ξ . Direct lending becomes an attractive alternative for the household only when intermediated investment is constrained and the deposit-loan spread is hence high. The idea behind this assumption is that banks have a superior screening and monitoring technology. This technological advantage explains why banks exist and is hence consistent with the fact that the financial sector intermediates a large part of investment in many economies⁷¹. Furthermore, allowing for an alternative, less efficient form of investment is in line with the finding that direct investment makes up for a significant part of the loss of intermediated investment during banking crisis, documented for example by Fiore and Uhlig (2015) for Europe and by Becker and Ivashina (2014) for the US. The cost ξ is allowed to vary with TFP. For positive $\hat{\xi}$ this dependence is such that the cost advantage of intermediated financing is particularly strong when the TFP is low. This reflects the idea that the screening and monitoring technology of the bank is most useful when profitable investment opportunities are scarce.

3 Financial frictions, sovereign debt, and the costs of default

This section discusses the consequences of the financial structure of the economy and its implications for the default decision.

3.1 Underinvestment due to the financial friction

In the model economy the household has two ways to investment into physical capital: either directly by lending to the firm, or indirectly by lending to the bank, which lends to the firm. (The bank also lends to the government, but abstract from that for the moment for the sake of simplicity.) Bank lending is assumed to be more efficient, because it comes free of any transaction costs, but the capacity of banks to intermediate is constrained by a leverage constraint. Whenever this constraint does not bind, competition assures that banks make no profits. But when it binds, it drives a wedge between the return on loans and the cost of deposits, even though banks behave competitively. At the same time the

⁷¹With the notable exception of the US, where market based funding plays an important role.

lower bound on dividends that banks have to respect implies a constant drain on the banks equity. To finance this drain without its equity being eroded, the bank constantly has to make profits. Assume the share η of pre-dividend equity that the banker can maximally retain is lower than the inverse of the gross return on loans. In this case, the return on the equity financed assets alone is not enough to finance this drain. Hence, for equity not to erode over time banks have to make profits also on deposit financed assets. These profits can only come from a spread between the deposit rate and the return on the banks assets, which in turn arises whenever the constraint binds.

Therefore, if η is small enough, as I assume, banks are *on average* constraint by both the deposit and the dividend constraint and the spread between the deposit rate and the return on assets is *on average* bigger than zero. This means that the return on intermediated investment that households perceive is lower than it actually is, whenever the constrain binds and *on average*. Households hence under-invest in intermediated loans. At the heart of this underinvestment in banks is an externality: Households only take the deposit rate into account, but not the profits of the bank, which flow back to the household lump sum. Even if they were not the recipient of these profits, they would value them in a non-atomistic setting, because they would understand that these profits allow banks to replenish their equity, which will reduce the tightness of the constraint and make them behave more competitively.

Furthermore, if the wedge or spread between the loan rate is high enough, households will find it optimal to invest directly loans, despite the costs associated with this transaction. This alternative channel of investment, which is absent from most papers on banking, has a counter intuitive effect. One might expect that it might dampen the effects of a shortage of bank equity, since it provides an alternative, albeit less efficient, channel of investment. Indeed, it does so. But at the same time it also makes the effects of the bank equity shortage worse, because it constrains the spread between deposit and loan rate above. This implies that the leverage constraint binds tighter and it slows down the re-accumulation of bank equity through retaining the profits derived from the lending spread.

3.2 A new motive for international borrowing

The fact that the household under-invests in physical capital has an interesting effect in the context of the full open economy model with a benevolent government. First note underinvestment means that even in the non-stochastic steady state the return on capital is higher than the inverse of the discount factor $1/\beta$. Second note that whenever the government borrows today to reduce taxes today, the household will want to save most of it, anticipating higher taxes tomorrow. These additional savings will, in normal times when the bank is not too tightly constraint, flow through the banking sector to the firms. They will earn a return equal to the deposit rate, which in the non-stochastic steady state

is equal to $1/\beta$, plus the spread sp . At the same time the funds have a cost of $1/\bar{q}$. Hence international lending is profitable from the benevolent government's point of view, as long as $1/\beta + sp \geq 1/\bar{q}$, that is as long as the domestic household is not much more patient than the rest of the world.

This means that the existence of capital, together with a friction that prevents the return on capital to converge to the world interest rate over time, constitutes a motive for government borrowing: The government is essentially arbitraging on the difference between the domestic and the world interest rate. Or, more loosely speaking, funds from international lenders are “cheaper” than taxes.

This return-differential motive complements the consumption-smoothing motive for participating in international financial markets.⁷² Yet, there is one important difference: The consumption-smoothing motive is roughly neutral across the cycle, in the sense that in good times the government wishes to save and in bad times to borrow. Conversely, the return-differential motive is not neutral across the cycle but on average calls for international borrowing, because for moderate deviations of TFP from its mean international funds are always cheap relative to domestic ones. Hence, this motive explains why the model economy is on average (and, for the calibrated version indeed most of the time) highly indebted.⁷³ While the return-differential motive for international borrowing is not new in general, the fact that the financial friction sustains this motive endogenously across the cycle is a novel feature of this model. It is interesting in particular in the context of the sovereign default literature, since most models of external sovereign debt have so far relied on the simplifying assumption that the borrowing country is strongly impatient (relative to the international lenders) to generate a substantial amount of debt.⁷⁴ As we will see in the quantitative part, this model works also without that assumption.

3.3 The costs of default

As I stressed before, the model I propose assumes no direct costs of default, neither the typical output costs nor exclusion. But this does not mean that default is costless. On the contrary, the model predicts sizable and long lasting costs of default.

Figure 1, which shows how the model evolves after a default in the absence of shocks,

⁷²Notice, that the country can smooth its consumption even in autarky by varying capital. Nevertheless, foreign investment has additional consumption-smoothing value, since its return is exogenous, unlike the return on capital.

⁷³Notice furthermore, that the cycle has opposing effects on the two motives: Low TFP constitutes a reason to borrow internationally from the point of view of the consumption smoothing motive. At the same time the return differential between foreign and domestic funds decreases with TFP and hence weakens the desirability of borrowing. The sum of the two effects, together with the varying costs of default determine how the amount and riskiness of government debt varies across the cycle.

⁷⁴With the time period being one quarter, Arellano (2008) uses a β of .953 and a net world interest rate of 1.7%, which implies a \bar{q} of 0.983. Aguiar and Gopinath (2006) use 0.8 and 1%. Mendoza and Yue (2012) use 0.88 and 1%. Joo (2014) and Gornemann (2015), which to the best of my knowledge are the only other papers which consider a sovereign default model with capital, rely on similar parameter values too.

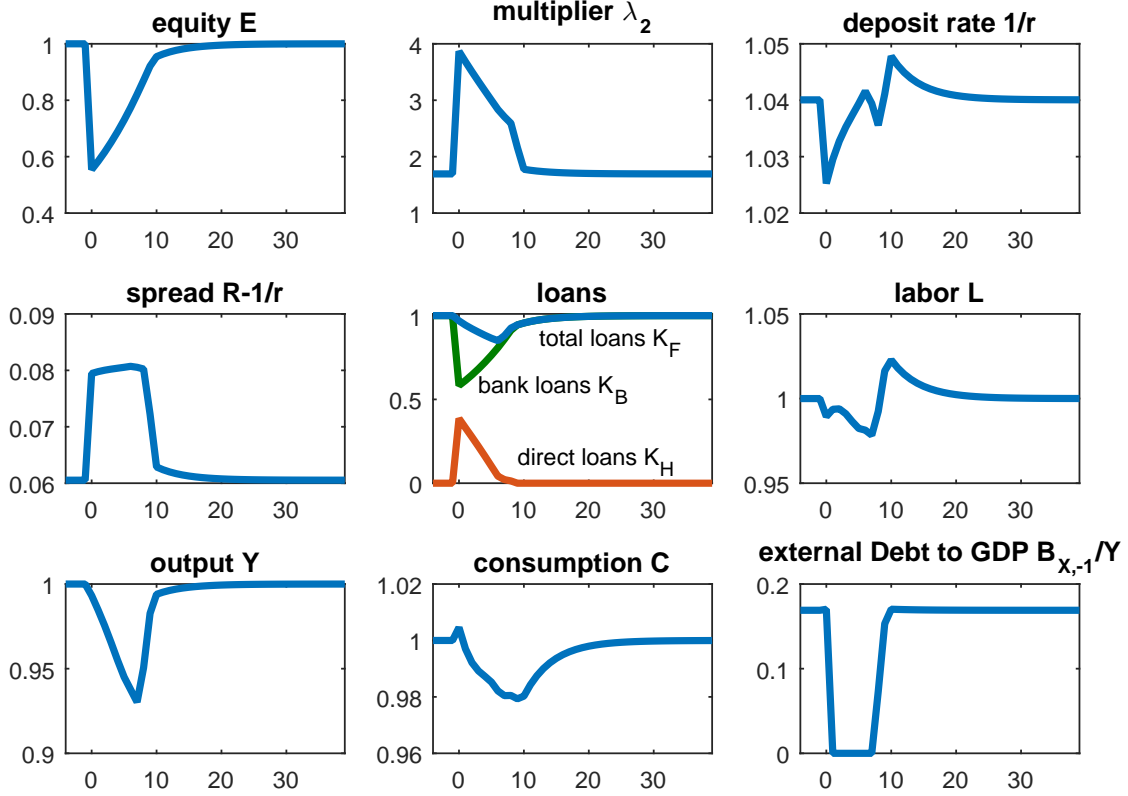


Figure 1: **Impulse responses to a default:** Prior to the default all variables are at their risk adjusted non-stochastic steady state. At period 0 the county defaults (note that this default is not optimal). No other shocks occur. Equity, capital, labor, output and consumption are normalized by their respective steady state values. Direct and bank loans are fractions of total capital. This picture is drawn for $\hat{\xi} = 0$ to make default at average TFP look more like TFP at low TFP.

illustrates the consequences of default. Since banks are exposed to sovereign debt, they loose a significant share of their equity. This implies that the leverage constraint, which has already been binding loosely before, now binds very tightly (see panel 2 of figure 1 for the multiplier of the leverage constraint). As a consequence, banks have to reduce the size of their balance sheets, and hence the amount of loans they extend and the amount of deposits they raise. The lower demand for deposits by banks makes the deposit rate drop, while the spread between the lending and the deposit rate increases. Given the high spread, direct investment becomes profitable for the household, despite the costs associated with this form of financing. Due to these costs however, the shift to direct financing can only partially make up for the reduction in bank financing and total loans, i.e. total capital drops. At the same time the household decreases his labor supply and his consumption⁷⁵. Since both labor and capital drop, this causes a fall in output. In the periods following default, banks slowly recapitalize by accumulating the profits generated

⁷⁵After a small initial jump in consumption due to the one-time gains from default. This jump is due to the fact that the figure plots the responses to default in the absence of shocks. In equilibrium default only occurs concurrent with bad shocks and the jump becomes negative.

by the higher spread on their lending activity. As they do so, they expand their balance sheet once again, and more and more of the direct investment is replaced by intermediated investment. The economy slowly converges back.

It is interesting to see what happens to external debt. After default has wiped out all foreign debt, the government does not borrow any new funds abroad. In fact they continue not to do so for several periods until they eventually enter the international capital markets again and borrow once more. What is happening? After default the government is *choosing* not to borrow any more from abroad. It does so because the main motive to borrow, which consisted in the difference between the world interest rate and the return on domestic capital, vanishes when financial intermediaries are so constrained, that the marginal return on capital is determined by the less efficient investment technology (direct investment). The government essentially self-excludes itself from the market for a period of several years.⁷⁶ The duration of the period is endogenous and depends on the duration of the financial crisis.

The model is hence able to endogenously explain not only the output costs of default but also exclusion from international capital markets. While other authors have provided models that endogenize the output costs of default, they still rely on the assumption of exclusion. In the case of Mendoza and Yue (2012) exclusion is even necessary to generate the output costs. This is the first model that can also account for exclusion endogenously. Furthermore, the mechanism provides a novel explanation for *why* we observe exclusion. The argument behind the assumption of exclusion typically refers to some coordinated punishment of the borrower by the lenders. Yet this argument has one weakness: as Kletzer (1994) argues, it would be optimal for individual (small) lenders to deviate from this collective punishment strategy. The explanation for exclusion this paper proposes is immune to such criticism because it is optimal for both sides not to trade. There is simply no price at which the borrowing country and the lender would want to trade. At the same time both this and the conventional theory are observationally equivalent.

In equilibrium all agents correctly anticipate the consequences of default: They understand that defaulting on foreigners per se is a free lunch. But they also understand that defaulting on local banks causes a financial crisis. If the benefits from default are low enough, i.e. if the external debt outstanding is low, the costs of a financial crisis outweigh the benefits of default. Hence *all* debt is repaid ex post, which makes *external* debt sustainable ex ante. The financial vulnerability of the economy essentially serves as a commitment device for a government that can't commit otherwise. Notice that this entails another externality: Banks do not get remunerated for the value of the commitment that their investment in bonds entails.

⁷⁶Notice that it might actually want to save abroad, but this is ruled out by assumption.

3.4 Default incentives

To understand what the default decision depends on it is useful to define the default set as those states where the government defaults in equilibrium:⁷⁷

$$\mathcal{D} = \{\Omega \in \Omega : V_H(Rep = 0|\Omega) > V_H(Rep = 1|\Omega)\}$$

This default set has a number of features that resonate similar findings in Eaton and Gersovitz (1981) and Arellano (2008). In particular proposition 2 and the first part of the conjecture are extensions of propositions in these papers, while the other parts of the conjecture are more specific to the context of this model.

Proposition 2 Assume that $\Omega_1 = [\omega_1, K_{H1}, D_1, B_{B1}, K_{B1}, R_1, B_{X1}] \in \mathcal{D}$. Then the following holds:

$$\Omega_2 = [\omega_1, K_{H1}, D_1, B_{B1}, K_{B1}, R_1, B_{X2}] \in \mathcal{D} \text{ if } B_{X2} \geq B_{X1}.$$

Proof : First, note that the value of defaulting $V_{GD}(\Omega)$ is independent of the level of B_X , hence $V_{GD}(\Omega_1) = V_{GD}(\Omega_2)$. Second, if $[T_2^*, \bar{B}_2'^*, X_2^*]$ denotes the optimal level of all choice variables under repayment given state Ω_2 , then $[T_2^*, \bar{B}_2'^*, X_2^*]$ is also feasible under repayment given state Ω_1 by the government budget constraint. Hence it must be that $V_{GR}(\Omega_1) \geq V_{GR}(\Omega_2)$. Summarizing, we have $V_{GD}(\Omega_1) = V_{GD}(\Omega_2) > V_{GR}(\Omega_1) \geq V_{GR}(\Omega_2)$, and hence that $\Omega_2 \in \mathcal{D}$. ■

Conjecture Assume that $\Omega_1 = [\omega_1, K_{H1}, D_1, B_{B1}, K_{B1}, R_1, B_{X1}] \in \mathcal{D}$. Then the following holds:

Part 1

$$\Omega_3 = [\omega_2, K_{H1}, D_1, B_{B1}, K_{B1}, R_1, B_{X1}] \in \mathcal{D} \text{ if } \omega_2 \leq \omega_1$$

Part 2

$$\Omega_4 = [\omega_1, K_{H1}, D_2, B_{B1}, K_{B1}, R_1, B_{X1}] \in \mathcal{D} \text{ if } D_2 \leq D_1$$

$$\Omega_5 = [\omega_1, K_{H2}, D_1, B_{B1}, K_{B2}, R_1, B_{X1}] \in \mathcal{D} \text{ if } K_{B2} \geq K_{B1} \text{ and } K_{H1} + K_{B1} = K_{B2} + K_{H2}$$

$$\Omega_6 = [\omega_1, K_{H1}, D_1, B_{B1}, K_{B1}, R_2, B_{X1}] \in \mathcal{D} \text{ if } R_2 \geq R_1$$

Part 3

$$\Omega_7 = [\omega_1, K_{H1}, D_2, B_{B2}, K_{B1}, R_1, B_{X1}] \in \mathcal{D} \text{ if } B_{B2} \leq B_{B1} \text{ and } B_{B1} - D_1 = B_{B2} - D_2$$

Proposition 3: Part 3 of the conjecture holds, if we allow the government to transfer funds from the bank's equity to the household in a lump sum fashion, but not the other way around.

⁷⁷Note that this definition does not exactly coincide with the definition in Arellano (2008), but has the same spirit.

Proof: First, note that conditional on default or repayment, the vector $[\omega, E, K_F, B_X]$ sufficiently summarizes the 7 dimensional state Ω (Here E denotes the pre-dividend pre-transfer bank equity $E = RK_B + RepB_B - D$). Second, note that the values of repayment are equal $V_{GR}(\Omega_1) = V_{GR}(\Omega_7)$ since the elements of the vector $[\omega, E, K_F, B_X]$ conditional on repayment are equal. Third, note that conditional on default, $[\omega, K_F, B_X]$ are equal but the banks pre-dividend pre-transfer equity is lower at Ω_1 . Denote the optimal level of all choice variables under default given state Ω_1 by $[T_1^*, \bar{B}_1^*, X_1^*, Z_1^*]$ where Z denotes the transfer from the bank to the household. Then, by choosing an appropriate transfer $Z_7 > Z_1$ the allocation $[T_1^*, \bar{B}_1^*, X_1^*]$ must be feasible given state Ω_7 . By optimality it must hence be that $V_{GD}(\Omega_7) \geq V_{GD}(\Omega_1)$. Hence $V_{GD}(\Omega_7) \geq V_{GD}(\Omega_1) \geq V_{GR}(\Omega_1) = V_{GR}(\Omega_7)$ and therefore $\Omega_7 \in \mathcal{D}$. ■

Proposition 2 and conjecture state that the default set is bounded below wrt. external debt B_X , domestic debt B_B , the fraction of bank-held loans to total loans $K_B/(K_B + K_H)$ and the loan rate R and bounded above wrt. TFP ω , deposits D and the ratio bonds over equity financed bonds $B_B/(B_B - D)$. The intuition behind proposition 2 is that if default is optimal for a given level of foreign debt it must be optimal for higher levels too, because the benefit of defaulting increases but the costs remain unchanged. The conjecture is intuitive too: Part 1 states that if default is optimal for a certain level of TFP, it must be optimal for lower levels of TFP too. This is so because TFP does not affect the bank equity, but it does affect the total resources of the economy. In worse states optimal savings are lower than in good states. Hence, bank equity, which is needed to intermediate savings, is less scarce at lower levels of TFP. Hence the loss of equity due to sovereign default can be expected to be less costly at lower TFP levels. At the same time aggregate resources are more scarce at lower TFP levels, hence the benefit of not repaying the rest of the world increases. Part 2 says that default is more attractive when the bank is better capitalized, keeping the total resources of the economy and the exposure of the bank fixed. Here the idea is that on one hand the cost of default, that is the loss of bank equity, diminishes as banks are better capitalized. At the same time the benefits of default increase as banks are better capitalized, since well capitalized banks allow a less distorted allocation of the fixed amount of total resources available. Part 3 finally states that lower levels of deposit financed sovereign debt exposure of the bank make default more attractive. This is straightforward given that the post default equity of the bank decreases in the bank's deposit financed exposure.

While the conjecture cannot be proven analytically, numerical experiments have consistently confirmed it.⁷⁸

⁷⁸Regarding part 1 of the conjecture: As the appendix shows, the numerical solution algorithm is based on the guess that part 1 holds. This guess is verified ex post at each grid point. Regarding part 3: The additional assumption necessary to proof part 3 of the conjecture does not change the equilibrium of the economy, if the value function of the government is decreasing in the share of total resources held by the bank, because then $Z^* = 0$. I find that in equilibrium the value function is decreasing across the whole state space for the calibrated model.

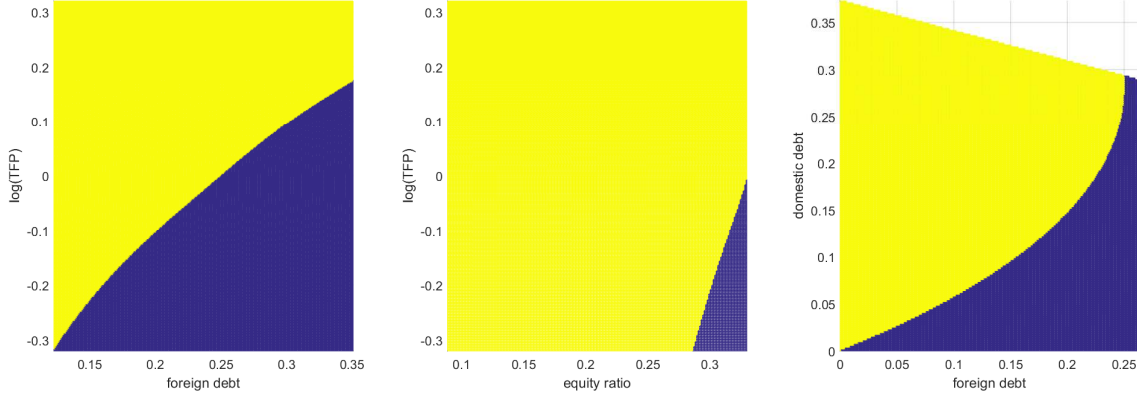


Figure 2: **Default Set:** The dark area is the default region, the bright area is the repayment region. The white area is off the equilibrium path. The remaining states are set to their non-stochastic risk-adjusted steady state values. The levels of foreign and domestic debt are normalized by the output at the non-stochastic risk adjusted steady state. The equity ratio in panel 2 is conditional on repayment.

Figure 2 shows the default set for the calibrated model, exemplifying the above characterizations. Panel 1 illustrates proposition 2 and part 1 of the conjecture: lower levels of TFP and higher levels of foreign debt make default more likely. Panel 2 illustrates the second part of the conjecture: the higher the ratio of bank equity to total resources of the economy (which may result from higher values of R , D or $K_B/(K_B + K_H)$) the more attractive default. Finally the last panel illustrates part 3 of the conjecture: The higher the domestic exposure to sovereign debt the less likely default.⁷⁹ This panel furthermore illustrates the disciplining role of domestic debt: the more exposed the domestic sector is, the more foreign debt will be repaid.

While the costs of default discussed above guarantee that some amount of external debt can be sustained, they do not suffice for default to happen along the equilibrium path. For default to happen, there needs to be furthermore a strong enough desire to borrow, such that the government has an ex-ante incentive to borrow even beyond the maximal amount, for which default is ex-post undesirable. Furthermore, the default costs need to depend on the exogenous state, such that it is possible to borrow an amount that ex post leads to default only for some of the possible realizations of the exogenous shock. While the motive for borrowing has been discussed before, the state dependence of the costs of default evident in panels 1 and 2 merits further explanation.

Figure 3 illustrates the dynamic difference in the net cost of default across the exogenous state. It plots the difference in the path of key variables for 2 scenarios, starting from the same initial conditions: In scenario (1) a TFP shock hits the economy in period 0 and the government repays. In scenario (2) the same TFP shock hits the economy but the

⁷⁹Note that here D is kept fix. That means the additional domestic debt is equity financed, so the graph actually exemplifies an even stronger conjecture that $\Omega_8 = [\omega_1, K_{H1}, D_1, B_{B2}, K_{B1}, R_1, B_{X1}] \in \mathcal{D}$ if $B_{B2} \geq B_{B1}$. For some pathological cases this stronger statement has been found to be wrong though.

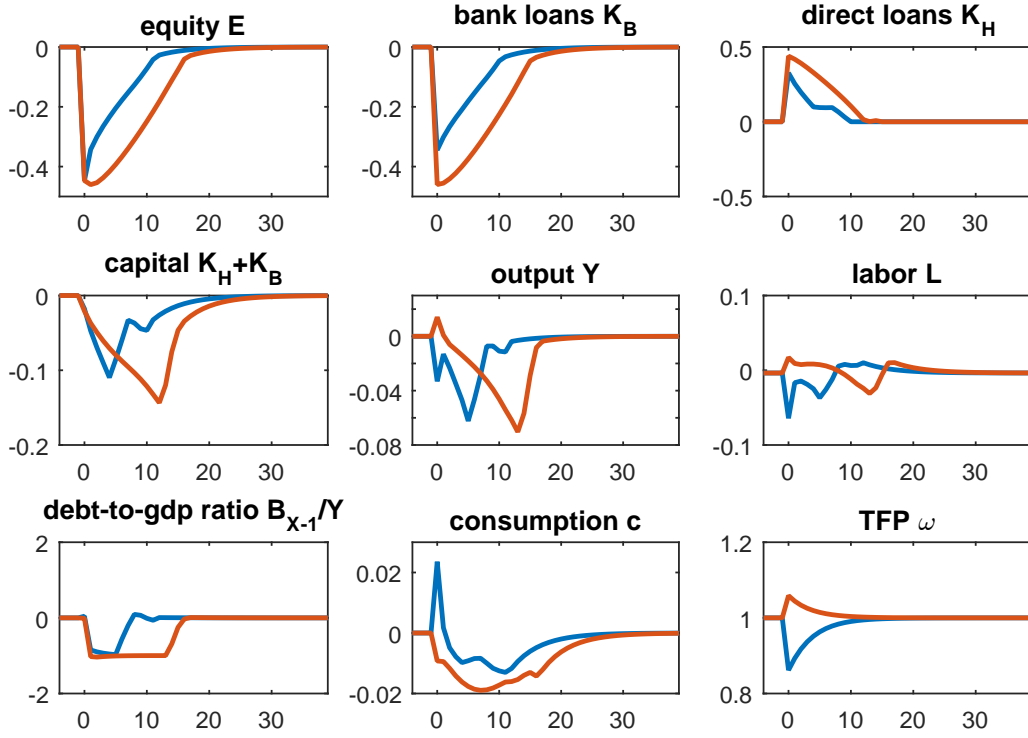


Figure 3: **Net costs of default:** This figure plots the difference between the paths associated with a two scenarios, both starting at the non-stochastic risk adjusted steady state. In scenario (1) in period 0 a TFP shock occurs and the government pays back. In scenario (2) the same TFP shock occurs at period 0 but the government defaults. No further shocks occur in either scenario. If the graph for e.g. output is negative, it hence means that (given the same TFP shock), output is lower under default than under repayment. The two lines correspond to a positive shock of one standard deviation (red line) and a negative shock big enough to trigger default (blue line). In case of the bad shock default is optimal; in case of the good shock repayment is optimal. All paths are normalized by their steady state values.

government defaults. The difference between these paths illustrates the net cost/benefit of default, which consists of the sum of the downside of default (banks' equity is eroded) and the upside (the foreign debt is not being repaid). This difference is plotted for two different realizations of the TFP shock: high and low. It is evident that the net costs of default in terms of consumption and labor (and hence utility) are much higher in case of the positive TFP shock. The intuition behind this is simple. In states of high TFP the households would like to invest more (both because expected returns are high and to smooth consumption), relative to the case of a negative shock. At the same time the loss banks face due to default does not depend on the exogenous state. Hence, the fall in bank equity has stronger distortionary effects when TFP is high. This can be seen in panel 4: The total amount of loans drops much more (relative to the repayment case) under the good shock than under the bad shock. These distortions constitute the cost of default. Given that the resource gain from defaulting on the rest of the world is also independent

of the state, but their utility value is higher in bad times, the net costs of default are bigger in states of high TFP.

Another subtle feature can be seen in this graph: Notice that the output costs of default in case of the bad shock, which is when default actually is optimal, increase over time. This is a result of the continued underinvestment, which results from the shortage of bank equity. As the economy under-invests in capital, the capital stock declines gradually (relative to the repayment case), which leads to a gradual decrease in output (relative to the repayment case). Hence the drop in output the results from a bad TFP shock and a simultaneous default is more persistent than the drop in output that would result from a bad TFP shock without default. This feature can help to explain the sluggishness of recoveries from default crisis.⁸⁰

4 Computation and calibration

4.1 Computation

Given the high complexity of the model, only an approximate solution of the government's problem (20) is obtainable. Solving a continuous DSGE model numerically typically requires the uses of 3 tools: function approximation, numerical integration and numerical maximization. I briefly summarize each of these three steps here. In doing so I highlight two computational innovations that were integrated in the solution algorithm. The appendix provides more details.

I approximate the policy and value functions $Z(\Omega)$ by twice continuously differentiable functions $\tilde{Z}(\Omega)$ over the whole state space (including exogenous states). In particular I use using cubic splines defined over a multidimensional Cartesian grid as recommended by Hatchondo et al. (2007). I extrapolate points outside the grid using these splines, but choose the grid such as to ensure that the maximum distance from the grid remains small and the probability of leaving the grid marginal.

Given this approximation for the policy and value functions, I *exactly* evaluate the integrals contained in the expectations of future values of these functions, such as $E_{|\Omega}(Z') = \int_0^\infty pdf(\omega'|\Omega)\tilde{Z}(\omega'|\Omega)d\omega$. This approach is novel and different from the quadrature approaches, that are usually used to evaluate expectations, including the application in Hatchondo et al. (2007). It extends or 'inverts' the insight of Gaussian quadrature methods, that integrals over polynomials can be evaluated exactly given a sufficient number of function values, to the piecewise-polynomial cubic-spline function. This approach has

⁸⁰Already the Arellano (2008) model has a similar feature: As the economy gradually recovers from a bad endowment shock, the default costs increase and keep endowment net of default costs depressed for longer than in case of repayment. Yet this is purely due to the recovery of the endowment. If the endowment were to stay constant for as many period as the economy remains in autarky, the default costs would remain constant too. In my model however the output costs of default increase after default, even if TFP were to remain constant.

two advantages: First, it avoids an additional layer of approximation. Second, and more importantly for the current application, it generates expectations that are twice continuously differentiable not only in the endogenous states, but also in the exogenous states. This facilitates the application of continuous solution methods to optimization problems where the exogenous state appears as a variable, as it for example does when there exists a threshold value of the exogenous state, such as in the case of default models the default threshold.^{81,82}

Finally, to find the stationary solution to the optimization problem of the government, given these choices for how to approximate policy and value functions and their expectations, I solve problem (20) recursively using a time iteration algorithm, that jointly iterates on the policy and value functions.⁸³ To find the solution at each individual time iteration, I use a numerical solver to solve the government's problem (20) continuously in all variables.

Solving the model poses some significant computational difficulties due to (1) the high dimensionality of the state space and (2) the complexity of the government maximization problem (20), which is subject to a system of nonlinear equations including complementarity conditions. To address the first problem, I use a trick that reduces the computationally necessary state space from 7 to 4 dimensions. The general idea is to replace the portfolio variables by cash-at-hand, which is only possible though after anticipating certain future decisions that depend on the whole state space.⁸⁴ This trick, which is the second methodological innovation, is explained in detail in the appendix and in its general applicability in Thaler (2016). Furthermore, the use of smooth interpolation and integration discussed above, allows me to operate with a low number of grid points without too much loss of precision. Second, to solve the complex maximization system, I rely on a modern, derivative based maximization algorithm for complementarity problems (KNITRO). Given I provide algebraically computed first derivatives of target and constraint functions and the smoothness of the optimization problem, this solver is able to reliably and quickly find solutions.

⁸¹Take for example the continuous version of the Arellano model, and be $\bar{\omega}$ the default output threshold and B the debt level. Then the expectations of the value function are given by $E_{|\Omega}(V(\omega'|B')) = \int_0^{\bar{\omega}} pdf(\omega'|\omega)V^{def}(\omega'|B')d\omega + \int_{\bar{\omega}}^{\infty} pdf(\omega'|\omega)V^{rep}(\omega'|B')d\omega$. Therefore $E_{|\Omega}(V(\omega'|B'))$ is a function not only of the endogenous state, B' but also of the exogenous state $\bar{\omega}$. Continuous interpolation of V implies continuity of the expectation in the endogenous state. If combined with exact integration it also implies continuity of the expectation wrt. the exogenous state.

⁸²This approach can be applied given any approximating function. Given a non twice-continuous approximation function, like a piecewise linear approximation, it is equivalent to straightforward quadrature, but then it does not deliver the second advantage. For chebychev polynomials, the other common twice continuous approximation function, it is equivalent to Gaussian quadrature with enough points.

⁸³First make a guess for the policy and value functions $c'(\Omega)$, $\lambda'_2(\Omega)$ and $V'_H(\Omega)$. Second, I solve the optimization problem for many points on a grid over the state space Ω . Third, I update my guess given the solutions from step two and then go back to step two.

⁸⁴Without anticipating future decisions, one can reduce the state space to 5 dimensions. Besides, by applying the same trick to the labor decision I could reduce the state by 1 further dimensions, which is left for future work.

4.2 Calibration

I calibrate the model to Greek data, which is the case that motivated this research. As we have seen, the model I laid out describes a small open economy whose government has access to international capital markets, with an important domestic financial sector, no foreign investment into the private sector and without the option to inflate away its debt. Greece satisfies this description to a large extent. First, Greece is a small open economy with respect to both the EU and the world. Second, the Greek government has borrowed extensively: the mean debt to GDP ratio for the post-Euro pre-default period 2000-2010 is 121% according to OECD data. This debt was held both domestically (48%) as well as abroad according to Bank of Greece data. Third, credit to the local economy is largely supplied by the domestic banking sector and non-intermediated investment plays a minor role (see e.g. Demirguc-Kunt and Levine (1999)). Similarly, domestic financial institutions account for most of the domestically held sovereign debt (80% according to Bank of Greece data). Fourth, private investment of foreigners in Greece play a minor role. As figure 4 shows, the net foreign asset position of Greece is dominated by government debt and what little net private investment there is largely flows out of Greece rather than in. Fifth, since the accession to the Euro zone, Greece can not inflate its debt away, even though they issued most of their debt under domestic currency (and law). The main difference between the model and the Greek scenario probably is the policy response: Both Greek banks as well as the government were to some extent bailed out by the EU. While the model abstracts from bailouts altogether, the bailouts that were implemented certainly did not go anywhere as far as to compensate entirely for the consequences of default.

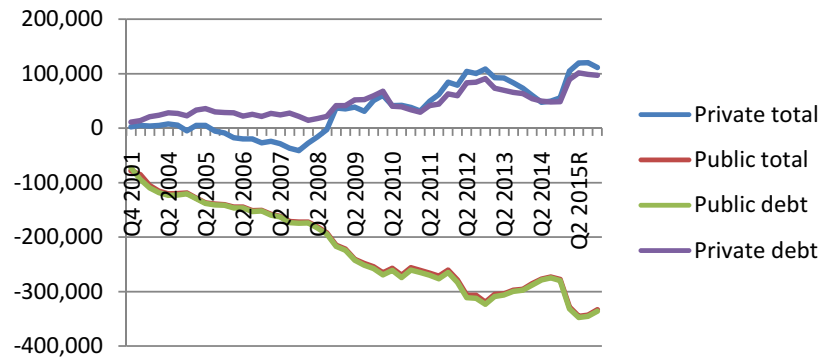


Figure 4: **International investment position:** Nominal values reported by Bank of Greece. The figure distinguishes between public and private liabilities, and between debt liabilities and total liabilities, including equity. For the government the latter two are almost identical.

While Greece seems an ideal match for the model, other southern European countries like Italy and Portugal also fit the above description. However in these cases default (resp. debt restructuring) did not materialize - even though it was a major concern at the height

of the European debt crisis - which makes them less useful examples for the following quantitative analysis. Spain and Cyprus could be considered similar examples as well, the latter even defaulted in the aftermath of the Greek debt restructuring. Yet for these two countries it was a financial crisis that triggered a sovereign crisis and not the other way around. While the modeling was guided by the European debt crisis, the predictions of the model furthermore are consistent with patterns documented by Gennaioli et al. (2014) across a large panel of mainly developing and emerging economies. Among these is the well studied case of the Argentinian default in 2001, where concerns regarding the damage to the financial sector indeed played a dominant role in the policy debate before and the policy response after default.⁸⁵ However, the mechanism the model describes can be expected to be stronger for countries with more developed financial systems.

The data on Greece used for the calibration and the model evaluation comes from several sources. Real GDP per capita is taken from the IMF IFS database. Its components are obtained from OECD data. TFP is taken from the AMECO database. Data on the balance sheet composition of Greek banks, the spread between short term household deposits and loans to the non-financial sector and the composition of Greek foreign liabilities is taken from the Bank of Greece. The sovereign spread is defined as the difference between the returns on one year government bonds from Greece and Germany. The data stems from the IFS for Greece and from the Bundesbank for Germany. Private credit is also taken from the IFS, while the sovereign debt to GDP ratio is taken from the OECD. For GDP and its components and TFP I use annual data covering the pre-default period 1980 to 2010, while for the financial variables the observation period is restricted to the post-Euro pre-default period 2000 to 2010. I do this both due to data availability and to account for the structural change of accessing the Euro, which should have a stronger impact on these variables. I decompose the series for GDP, TFP and credit into a linear trend and a cyclical component and use the deviation of each variable from its trend. Note that in doing so I follow Arellano (2008), who also uses a linear long-run trend, whereas many other papers using emerging economy data focus on the business cycle frequency fluctuations alone (E.g. Mendoza and Yue (2012), Aguiar and Gopinath (2006)). While these differences are relatively unimportant for the much studied 2001 default of Argentina, they are significant for Greece, which experienced a much more prolonged and deep downturn, which a medium-frequency filter would attribute to the trend. Furthermore, using data only till 2010 to construct the trend, (which I then extrapolate for the post default years), prevents the estimate of the trend to be affected by the 2010 crisis. This choice is in line with my model, in which sovereign defaults are rare events that lead to large and prolonged temporary downward deviations from the trend. The trade balance is expressed in percentage of GDP.

The calibration is summarized in table 1. As in the data the time period is one year.

⁸⁵ See Sosa-Padilla (2012), Perry and Servén (2003) and Kumhof (2004).

Parameter	Value	Target/Explanation
α capital share	0.36	standard value
δ depreciation	0.14	21% investment/GDP
β discount factor	0.96	same as β
γ risk aversion	2	standard value
χ labor weight in utility	1.33	mean of labor = 1
ν inverse Frisch elasticity	0.5	standard value
G government expenditures	0.28	21% govt. consumption/GDP
θ financial constraint	0.7	28% equity ratio
η share of retained equity	0.803	6% deposit-loan spread
ψ share of bonds on balance sheet	0.14	14% exposure
$\bar{\xi}$ average cost of direct investment	0.071	1.5 ppt increase of deposit-loan
$\hat{\xi}$ dependence of cost of direct investment on GDP	0.45	spread in case of default, costs vary 2 ppt across the cycle
\bar{q} inverse world interest rate	0.96	real rate Germany
ρ persistence of TFP shock	0.76	estimate
σ standard deviation of TFP shock	0.057	7.5% standard deviation of GDP

Table 1: **Parameter values**

The utility function features standard parameter values: The discount factor β is set to 0.96 while the risk aversion coefficient γ is chosen to be 2, a common value in the default literature. The labor weight χ is set to normalize the mean of labor to 1 while ν is chosen to imply a Frisch elasticity of labor of 0.5 as in Boccola (2016), which is a common value and within the range of empirical estimates reported by Greenwood et al. (1988). The production function is parametrized in with a conventional capital share α of 36%. Depreciation δ is set such as to match the 21% average investment to GDP ratio for Greece. G is chosen to match the average government consumption to GDP ratio of 21%.

The inverse world interest rate \bar{q} is set to 0.96, that is to exactly the same value as the domestic discount factor. This value is in line with the average real interest rate in (West) Germany for the period 1980 to 2010. While this choice might not seem very surprising, it is highly unusual in the context of quantitative models of sovereign default, which usually rely on a huge difference between β and \bar{q} in order to generate a motive for the government to borrow. As I explained before, my model features an endogenous motive for borrowing and can therefore do without this assumption.

The choice of the financial sector parameters is the least straightforward. I choose these values in order to match moments of financial variables for Greece. Given the previous parameters, the payout parameter η is chosen to match an average spread between the loan and the deposit rate of 6%. This is in line with the average spread between 1 year deposits and firm loans for Greece. The parameters that govern the relative cost

advantage of intermediated over direct investment $\bar{\xi}$ and $\hat{\xi}$ are simultaneously chosen such that direct investment becomes profitable only after an extreme shock (like default) and so as to match the 1.5ppt increase in the deposit-loan spread after default observed in the Greek data. This implies that ξ varies only mildly.⁸⁶ The share of domestic bonds on banks balance sheets ψ is set to 14%, which is the mean of the Greek credit institutions' exposure to sovereign bonds in the Bank of Greece data. Lastly, θ , which determines the tightness of the banks' (implicit) leverage constraint, is set to 0.7, which implies a mean equity ratio of 26% percent. This value, while far higher than the average for Greek banks⁸⁷, is approximately in line with the well known papers by Gertler and Kiyotaki (2010) and Gertler and Karadi (2011), which feature a leverage constraint that is very similar in spirit. Furthermore, this value guarantees that sovereign default does not drive the banks into default in the ergodic distribution. Hence the higher-than-realistic equity ratio compensates for the simplifying assumptions that default is complete and there is no partial recapitalization of banks after default.

Finally, the the persistence parameter of the TFP process ρ is estimated using detrended TFP data. The corresponding standard deviation σ is set such as to match the standard deviation of detrended GDP in the data.⁸⁸ To compute the volatility of GDP for the model, following the literature, I simulate the model for 100.000 periods, extract the 31 periods prior to each default, express the data in deviation from its mean and report the mean of the volatilities of the pre-default periods.

5 Quantitative results

This section assesses the model's quantitative performance in terms of business cycle moments and the patterns observed around default episodes.

5.1 Cyclical moments

Table 2 reports the most important business cycle moments of the model and compares them to their empirical counterparts in Greek data.

The data used to generate the empirical moments has been discussed before, with the exception of the default frequency. This frequency is an empirically difficult concept: Since default is an extremely rare event, it is very hard to estimate this value well given a relatively short observation period. If one considers the longest possible observation

⁸⁶ ξ moves from 0.06 to 0.08 as log TFP moves from 3.5 standard deviations above to 3.5 standard deviations below the mean.

⁸⁷In the period 2000-2009 the average equity ratio for Greek credit institutions was 8.3%. In the aftermath of the crisis Greek banks have massively expanded their equity buffers, and have now reached an equity ratio of 21.3%.

⁸⁸As commonly found in the RBC literature (e.g. Cooley and Prescott (1995)), the estimated volatility of TFP is too small for the model to explain observed volatilities. Hence, to capture the observed volatility of output in a simple model with just one shock I scale up the estimate of σ .

Moment		Data	Model
Means			
$E(\bar{B}/Y)$	Total debt to GDP ratio	121%	44%
$E(B_X/Y)$	Foreign debt to GDP ratio	51%	18%
$E(B_X/\bar{B})$	Share of domestic debt	47%	59%
$E(spread)$	Sovereign spread	0.48%	0.50%
$E(Rep)$	Default frequency	< 2.7%	0.44%
Volatilities			
$\sigma(Y)$	GDP	7.5%	7.4%
$\sigma(\bar{B}/Y)$	Total debt to GDP ratio	7.3%	4.2%
$\sigma(TB)$	Trade balance	3.7%	2.9%
$\sigma(spread)$	Sovereign spread	0.55%	0.26%
$\sigma(C)/\sigma(Y)$	Relative consumption volatility	79%	83%
$\sigma(I)/\sigma(Y)$	Relative investment volatility	226%	306%
Correlations of GDP with			
$corr(Y, C)$	Consumption	95%	93%
$corr(Y, I)$	Investment	91%	81%
$corr(Y, spread)$	Sovereign spread	-28%	-70%
$corr(Y, TB)$	Trade balance	11%	-1%

Table 2: **Cyclical moments**

period, that is since Greece's independence in 1829, there have been 5 default events: The 4 reported by Reinhart and Rogoff (2008) in 1843, 1860, 1893, 1932 plus the recent crisis. This yields an annual default frequency of 2.7%, which is in line with similar estimates in the literature for groups of emerging economies over the last centuries. Yet, this estimate is based on the assumption that the underlying economy has not changed over time. This is hardly the case for Greece. Financial development for example arguably has made default more costly. Given that my model describes a financially developed economy, in which the costs of defaults are presumably larger than they were in past centuries or are in economies with less developed financial markets, I consider this estimate an upper bound for the default frequency at best. The low spreads on Greek bonds in the post-Euro pre-default period seem to confirm this notion.

To generate the moments for the stochastic stationary state of the model, I simulate the model for 100,000 periods, discarding the initial observations. From this series I calculate the default frequency. Then I take the values for the 31 periods preceding each default event (excluding cases where another default happened in this period), normalize them by their mean and calculate the respective moments for each pre-default episode. Table 2 reports the means of these moments. Notice that, with the exception of the output volatility, none of these moments were targeted.

The model predicts a total debt to GDP ratio of 44% and a foreign debt to GDP

ratio of 18%, which is less than half of their empirical counterparts. This is a well known property of models with short term debt. For example, Arellano (2008) and Mendoza and Yue (2012) report even lower numbers: They obtain 6% and 23% quarterly debt to quarterly GDP, or 1.5% and 5.75% debt to annual GDP. As Chatterjee and Eyigungor (2012) and Hatchondo and Martinez (2009) show, more realistic ratios can be obtained with long term debt, yet this extension is beyond the scope of this paper. The magnitude of the volatility of the debt to GDP ratio lies in the ballpark of what is measured in the data⁸⁹. Furthermore the model predicts that 59% of debt are domestically held, which is not very far away from the 47% observed in the data.

Like in all models with risk neutral lenders, the default frequency and the mean of the sovereign spread coincide by assumption (almost⁹⁰). Hence, the model can only match one of the two numbers. In fact the model matches the spread very well, which has the consequence that the predicted default frequency is much lower than the “historical upper bound estimate” discussed above. Furthermore the model yields a negative correlation between GDP and the spread, and in doing so reproduces a stylized fact that holds not only for the Greek default episode studied here, but is generally found in studies of sovereign debt (see e.g. Mendoza and Yue (2012)). At the same time the model underestimates the volatility of the spreads.

Besides, the model preserves the typical features of the real business cycle model: As in the data, consumption is slightly less volatile and investment significantly more volatile than output. Moreover, the model matches the data well in predicting a very high correlation between consumption and GDP, and, albeit less perfectly, a slight lower but still high correlation between investment and output. Furthermore the model predicts a trade balance that is about as volatile and as uncorrelated to GDP as in the data.

5.2 Default episodes

Next I study the model’s dynamics around default events. This is done using event study techniques. Similarly as before, I extract the period starting 5 years before default and ending 25 years after from the simulated data and normalize these values by their respective means. I plot the mean of the evolution of each variable in figure 5. To highlight the stochastic properties of these paths, their 10th to 90th percentiles are also plotted. The figure furthermore shows the evolution of the respective variables in the data. The default period 0 corresponds to 2011 for the data.⁹¹ GDP values beyond 2015 are IMF

⁸⁹Note that this number is very poorly estimated. Excluding the last observation (2010) slashes this number by 2/3

⁹⁰The following relationship holds $spread = \frac{P(Rep=0)}{1-P(Rep=0)} \frac{1}{\bar{q}}$, which for \bar{q} and $1 - P(Rep = 0)$ close to 1 can be approximated as $spread \simeq P(Rep = 0)$. Furthermore, the default frequency refers to the whole sample and the spread to pre-default periods.

⁹¹The default of Greece was a somewhat gradual process. While Greece effectively had lost access to capital markets by the end of 2010, Greek banks registered their first losses on sovereign bonds in August 2011 when they participated in the securities exchange program. The process of debt restructuring was

forecasts.

Several features are worth noticing. First, in line with the other two stylized facts documented by Mendoza and Yue (2012) the model predicts that default happens in bad times. Prior to default output is below trend 2/3 of the time and default itself is associated with a strong drop in GDP. Furthermore the external debt ratio is above average prior to default and peaks in the default period. Second, the magnitude of the output drop associated with default predicted by the model coincides with the what we observed in Greece, where GDP dropped by a quarter after default. Third, the model predicts a credit crunch: Funds supplied to the real economy drop, while the spread between the deposit and loan rate spikes. Again, the magnitudes predicted by the model are similar to what we observe in the data. Fourth, while the model understates the magnitude of the drop in consumption, it does a fairly good job with respect to investment. Fifth, the responses predicted by the model are less protracted than what we observe in Greece, despite the fact, that the output costs increase over time (recall figure 3). This divergence is explained by the relatively quick reversal of TFP in the model and the linear detrending applied to the data.

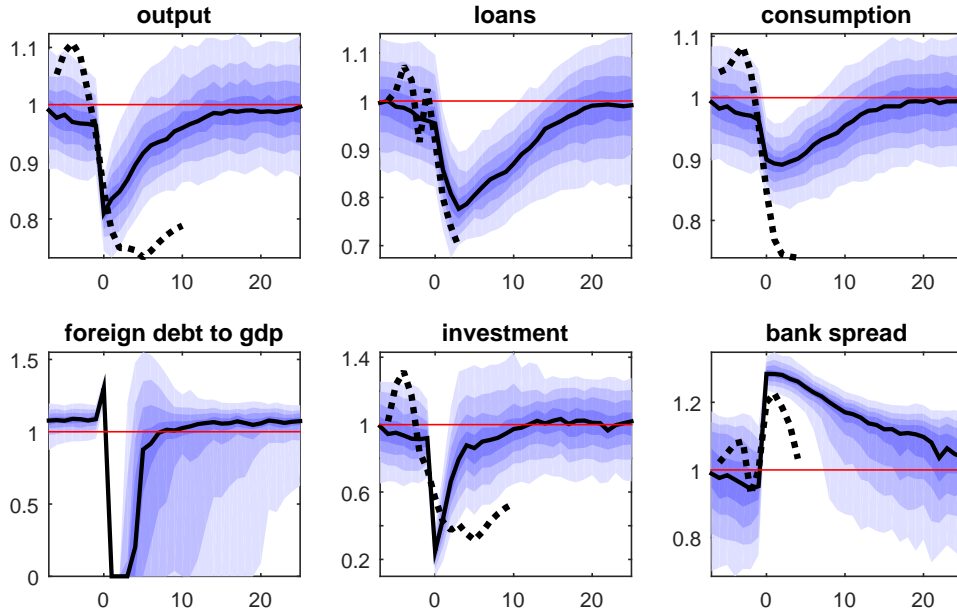


Figure 5: **Default episodes:** This figure plots the mean of the simulated path of key variables around default events (black solid line). Furthermore the 10th to 90th percentiles are plotted in blue. The black dotted line corresponds to the evolution of the variable in Greek data, where loans are equated to private credit. All variables are normalized by their means. Default occurs at period 0. For the data this period corresponds to 2011.

Most importantly, the model predicts that default is followed by a period of several years, during which the government does not issue any new debt. Unlike in existing finished in April 2012. When mapping this process to the more stylized default decision in the model, it seems reasonable to choose 2011 as the default period.

models of sovereign default, this is an equilibrium outcome. The average duration of this period is 4 years, which is close to the values usually assumed in the default literature. With Greece still excluded from the capital markets as of summer 2016, we still have to see how long it will take till Greece re-accesses the markets. Yet the prediction of the model is well within the range of 1 to 8 years estimated in Gelos et al. (2011), Richmond and Dias (2009) and Cruces and Trebesch (2013). Furthermore, note that the endogenous duration of exclusion is stochastic. As figure 6 shows, its distribution is strongly skewed to the right, with a mean of 5.8 and a median of 4 years.⁹² This property of the model is in accordance with empirical observations: Using a sample of 147 defaults since 1975 Schmitt-Grohe and Uribe (2016) report a mean duration of default of 8 and a median of 5 years. However, their sample contains a lot of developing countries. Constraining the sample to the 50% more financially developed countries according to the financial institutions development index proposed by Svirydzhenka (2016)⁹³ reduces these values to 3 and 6.⁹⁴

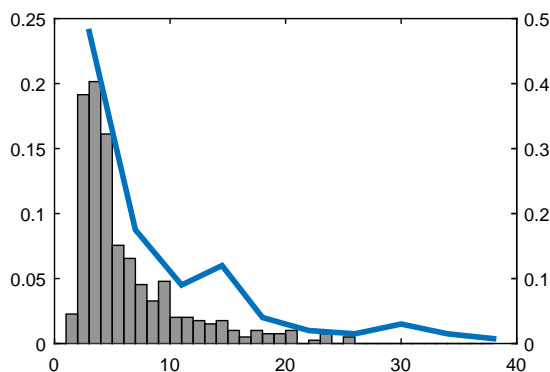


Figure 6: **Length of default episodes:** The bar histogram refers to the model, the distribution function (blue line, right axis) is copied from Schmitt-Grohe and Uribe (2016), Figure 13.1.

⁹²I consider a country to have re-accessed capital markets if it borrows for two consecutive periods. This way, the exclusion periods include a few episodes where the government interrupts its no-borrowing period for one period due to a strong temporary shock.

⁹³Thereby I exclude defaults of most African and less developed Latin countries.

⁹⁴Two caveats are in order: (1) Schmitt-Grohe and Uribe (2016) report the duration of default, not of non-zero trade, which need not necessarily coincide. (2) There are several plausible ways to define the start and end dates of default and market access which lead to deviations across different studies. However, the skewness seems to be a robust feature.

Gelos et al. (2011) measure the time from default to reaccess and report a mean of 4.7 and a median of 3.5 period across 45 default episodes. Combining the data from Schmitt-Grohe and Uribe (2016) to date entry into default and Cruces and Trebesch (2013) to date reaccess to the markets and restricting the set to the more developed countries leaves us with only 38 episodes and yields a median of 7.5 and a mean of 9. Among the episodes I referred to before as cases where the financial sector played an important role, the exclusion period is significantly shorter: Greece (2012) 4+, Cyprus (2013) 2, Argentina (2001) 5, Russia (1999) 2.

5.3 Robustness

The results are qualitatively robust against a number of qualitative changes. The exact shape of the financial friction for example does not influence the results much. In particular, using a state dependent leverage constraint as in Gertler and Kiyotaki (2010) or a simple leverage constraint, does not affect the results much

⁹⁵. Furthermore, the main results are robust to using GHH preferences or abstracting from labor all together. Using government spending / endowment shocks produces similar results as well.

Furthermore, most of the conventional parameters have relatively little quantitative effect on the default statistics of the model as well. This is in particular true for the level of government spending and the parameters of the utility function and the production function. Yet, allowing the domestic economy to be impatient relative to the world interest rate, as the literature routinely does, the model yields significantly higher default rates.

The financial sector parameters play a more important role, since they determine the costs of default. The parameter ψ , which determines the exposure of banks to sovereign debt and hence the strength of the financial crisis triggered by default, increases the costs of default and hence the amount of debt which is sustainable. The intensity of the financial friction θ determines the leverage of banks. The higher the leverage, the higher the costs of default and hence the amount of sustainable debt. The dividend payout parameter η pins down the spread between the bank lending and the deposit rate. The higher the spread, the higher the difference between the steady state return on domestic capital and the world rate, hence the stronger the incentive for the government to borrow - more and at higher risk. Finally the cost advantage of intermediated investment ξ has an ambiguous effect on the default incentives. Consider first the case that ξ is constant, i.e. $\hat{\xi} = 0$. While within one period, a lower cost advantage of intermediation means that the reduced intermediation capacity of banks due to the shortage of bank equity is less harmful for the capital stock, it also means that the maximum spread which banks can charge is lower. This slows down the process of bank equity reaccumulation, and extends the period, during which bank equity is so scarce that direct intermediation is used. For the calibration I use I found that the second effect dominates the first (locally), and hence

⁹⁵A simple leverage constraint would require the bank to finance no more than a certain fraction of their assets by deposits: $d'r \leq \theta(k' + qb')$. As in Iacoviello (2005) I assume that the bank is required to promise no more than a certain fraction of their net return to depositors $E_{|\Omega} [u_c(c')]d' \leq \theta E_{|\Omega} [u_c(c')(Rep' b' + R' k')]$ Gertler and Kiyotaki (2010) assume that the continuation value of banking has to be bigger than a certain (divertable) fraction of the assets $E_{|\Omega} [V'_B] \geq \theta(k' + q'b)$. These three different constraints have slightly different implications: For a simple leverage constraint interest rates do not matter, and book equity is constant as long as the constraint binds. For the constraint I use today's interest rates matter. The higher the spread between the loan and the deposit rate, the higher today's leverage is allowed to be. This is so because higher bank profits tomorrow increase the value of the bank in case it is taken over by depositors. The constrained proposed by Gertler and Kiyotaki (2010) extends this reasoning to the infinite future: the discounted value of the profits in all future periods disciplines the banker. Another way to look at the three constraints is to say that the first constrains the book-value leverage, the third constrains the market value leverage and the second lies in between.

lowering ξ marginally increases both the period of zero external debt and the sustainable amount of debt. That lower ξ increases the debt capacity only locally becomes clear when one takes the value for ξ towards its lower extreme of zero. Trivially, at $\xi = 0$ banks become obsolete and no debt is sustainable. The other extreme, setting ξ to infinity, means shutting down direct investment and turns out to have similar effects: in that case upon default the economy experiences a very strong and very short financial crisis, during which investment drops dramatically and the deposit-loan spread shoots up to empirically very implausible levels, which though allow the bank to recapitalize almost fully within one period. The cost of this strong but short crisis is far smaller than the cumulative cost of the protracted crisis resulting from intermediate values of ξ and hence the amount of debt sustainable is lower. This highlights the crucial role of allowing for an imperfect substitute for bank intermediation to generate bigger and more protracted financial crisis. Allowing ξ to vary slightly with TFP makes the post-default bank lending spread and recovery path depend on TFP in such a way that, sovereign default becomes more attractive in bad times (in relative terms), and hence slightly increases the default probability (by about 0.1 ppt) and the debt to GDP ratio (by 1pt) and the exclusion time (1 period).

6 Policy application: The EU plan to reduce the exposure of banks to domestic sovereign debt

In the aftermath of the sovereign debt crisis, policy makers in the EU have begun discussing new rules to weaken the sovereign-bank nexus. One proposal being discussed is to limit the exposure of banks to domestic sovereign debt through various measures like ceilings or higher capital requirements on domestic sovereign exposure.⁹⁶ The intention of this proposal is to reduce the risk of a vicious cycle, in which problems of sovereign debt sustainability lead to losses for banks, which in turn affects output and hence government revenues negatively, and hence makes the sovereign debt burden even more unsustainable. Yet, looking at this proposal through the lens of this model, such a reform may have unintended consequences: Making banks less exposed to domestic sovereign, *ceteris paribus*, debt reduces the cost of default for the domestic economy. This has two implications. First, default will be more attractive *ex post*, given the same amount of external debt. Changing the rules from one day to the other may hence trigger sovereign defaults that would otherwise not have occurred. Second, the anticipation of lower default by international lenders *ex ante* costs means that governments will not be able to borrow as much as before. Hence, the benefits derived from international lending will shrink.⁹⁷

⁹⁶See for example “Sovereign debt rule changes threaten EU bank finances”, Financial Times, 8.6.2016

⁹⁷In the policy discussion it is sometimes argued that the reduced demand for government bonds by domestic banks may increase the interest rates of sovereign bonds as an unintended side effect. Yet

Moment		Baseline	ψ -20%
$E(\bar{B}/Y)$	Total debt to GDP ratio	43%	29%
$E(B_X/Y)$	Foreign debt to GDP ratio	18%	9%
$E(B_X/\bar{B})$	Share of domestic debt	60%	69%
$E(spread)$	Sovereign spread	0.47%	0.65%
$E(Rep)$	Default frequency	0.44%	0.62%

Table 3: **Effect of reduction of exposure**

To illustrate these consequences, I analyze the effects of a reduction of the banks exposure to sovereign debt ψ by 20%.⁹⁸ Table 3⁹⁹ shows the long run effects of this regulatory change: As the banks exposure is reduced by 20%, the total debt to GDP ratio drops by one third and the foreign debt ratio by half. This loss of foreign debt capacity comes at a utility cost of about 0.9% in consumption equivalent.

The short run effect is equally stark: Assume the country finds itself at the non-stochastic risk adjusted steady state of the baseline economy at the end of period t . Assume furthermore that between periods t and $t+1$ ψ is changed unexpectedly, but the balance sheet of the bank is not yet affected. Then there is a 10% probability that TFP turns out so bad that the government will default at the beginning of $t+1$ because it can only partially roll over its outstanding foreign debt and prefers default to the high taxes otherwise needed. If we assume instead that the change in the regulation is affecting also the current balance sheet between the periods¹⁰⁰, the effect is even more striking: As the government wakes up the next morning, not only does it realize that it can't roll over its debt, but also that its banks are less exposed. It will default with 86% probability. While these sudden change scenarios are arguably too stylized, it exemplifies the point that if the default probability along the transition to the lower ψ is to be contained, a gradual implementation of the reform is to be recommended.¹⁰¹

7 Conclusion

Motivated by the recent European sovereign debt crisis, this paper proposes a quantitative model of opportunistic sovereign default with endogenous costs of default. In particular, it

this argument is simply based on lower *demand*, while the argument I make here is based on the lower *commitment value*. In fact the model abstracts from the demand argument altogether: In the model domestic borrowing by the government does not add any resources to the economy and taxation is lump sum, so lower domestic demand per se is irrelevant. Domestic borrowing is useful only because it provides the commitment to borrow externally.

⁹⁸This number is arbitrary, as there is no concrete final proposal yet.

⁹⁹Unlike in the previous table, these moments are unconditional.

¹⁰⁰Assume that 0.2% of the government debt on the domestic bank's balance sheet is exchanged with the household for a claim of equal amount.

¹⁰¹It seems likely that a model with long term debt would predict even starker consequences.

aims at endogenizing two types of default costs: the lack of access to international capital markets and output costs. With this objective, I develop a model of sovereign debt with international lenders and an explicitly modeled domestic economy, which consists of households, banks and firms. Savings are intermediated from households to firms by banks. Since there is a friction in the intermediation process, which makes bank equity both necessary and scarce, the economy never reaches its optimal level of investment. This fact generates a strong incentive for the government to borrow on international markets. While the government might default on this debt ex post, the fact that domestic banks are also exposed to sovereign debt generates commitment: If the government chooses to default, banks suffer losses, which leads to a period of credit shortage. This translates into a reduced capital stock and output losses. Furthermore, while the financial crisis lasts, the domestic return on investment drops which discourages the government to borrow at the constant world interest rate. Hence default is followed not only by a period of depressed output, but also by a period during which the government issues no new bonds. These two consequences of default arise endogenously and make some amount of debt sustainable.

Conceptually, this paper therefore contributes to the existing literature in three ways. First, it introduces the capital shortage motive for borrowing into the sovereign debt literature. Second, it proposes a new mechanism to explain the output costs of default, which is in line with recent experiences with the sovereign bank nexus. Third, it proposes a novel explanation for why we observe temporary breakdowns of international borrowing in the aftermath of default and for how long this period lasts, which is a question that has not been addressed by this literature before.

Quantitatively I show that the model calibrated to other moments predicts empirically plausible magnitudes of the consequences of default: Both the drop in output, investment and credit as well as the duration of market shutdown are roughly in line with the data.

This model furthermore offers insights that are relevant for a current policy debate on reducing banks exposure to domestic sovereign debt. I show that any such policy not only makes banks more resilient against sovereign debt crisis, which would be desirable per se. It also reduces the sustainability of sovereign debt, which is clearly undesirable, at least in the context of the model. Furthermore this model also contributes an explanation to why some countries like the US, Japan, or Italy seem to be able to sustain debt to GDP ratios that are higher than what other countries can sustain. It suggests that countries which have a higher share of domestically held debt and which have more leveraged financial sectors would suffer more in case of default and are hence less likely to renege on their obligations. The same argument may apply over time: The development of leveraged financial sectors across the developed world may explain the increase of debt to GDP ratios in the post war period.

References

- Acharya, V. V., Drechsler, I., and Schnabl, P. (2011). A Pyrrhic Victory? - Bank Bailouts and Sovereign Credit Risk. NBER Working Paper 17136, National Bureau of Economic Research, Inc.
- Aguiar, M. and Gopinath, G. (2006). Defaultable debt, interest rates and the current account. *Journal of International Economics*, 69(1):64–83.
- Arellano, C. (2008). Default Risk and Income Fluctuations in Emerging Economies. *American Economic Review*, 98(3):690–712.
- Basu, S. (2010). Sovereign Debt and Domestic Economic Fragility.
- Becker, B. and Ivashina, V. (2014). Cyclicalities of credit supply: Firm level evidence. *Journal of Monetary Economics*, 62:76–93.
- Boccola, L. (2016). The Pass-Through of Sovereign Risk. *Journal of Political Economy*.
- Bolton, P. and Jeanne, O. (2011). Sovereign Default Risk and Bank Fragility in Financially Integrated Economies. NBER Working Paper 16899, National Bureau of Economic Research, Inc.
- Broner, F., Martin, A., and Ventura, J. (2010). Sovereign Risk and Secondary Markets. *American Economic Review*, 100(4):1523–55.
- Brutti, F. (2011). Sovereign defaults and liquidity crises. *Journal of International Economics*, 84(1):65–72.
- Chari, V., Dovis, A., and Kehoe, P. (2015). On the Optimality of Financial Repression. Working Paper.
- Chatterjee, S. and Eyigungor, B. (2012). Maturity, Indebtedness, and Default Risk. *American Economic Review*, 102(6):2674–2699.
- Cooley, T. F. and Prescott, E. C. (1995). Growth and business cycles. In Cooley, T. F., editor, *Frontiers of business cycle research*, pages 1–38. Princeton University Press, Princeton.
- Cruces, J. J. and Trebesch, C. (2013). Sovereign Defaults: The Price of Haircuts. *American Economic Journal: Macroeconomics*, 5(3):85–117.
- De Paoli, B., Hoggarth, G., and Saporta, V. (2006). Costs of sovereign default. Technical Report 03/06.

- Demirguc-Kunt, A. and Levine, R. (1999). *Bank-Based and Market-Based Financial Systems: Cross-Country Comparisons*. Policy Research Working Papers. The World Bank.
- D’Erasmus, P. and Mendoza, E. (2012). Domestic Sovereign Default as Optimal Redistributive Policy. University of Maryland Working Paper.
- Eaton, J. and Gersovitz, M. (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis. *Review of Economic Studies*, 48(2):289–309.
- Engler, P. and Grosse Steffen, C. (2013). A Theory of Sovereign Risk Premia and Financial Intermediation in the Euro Area. DIW working Paper.
- Engler, P. and Grosse Steffen, C. (2016). Sovereign risk, interbank freezes, and aggregate fluctuations. *European Economic Review*, 87:34–61.
- Erce, A. (2012). Selective sovereign defaults. Globalization and Monetary Policy Institute Working Paper 127, Federal Reserve Bank of Dallas.
- Fiore, F. D. and Uhlig, H. (2015). Corporate Debt Structure and the Financial Crisis. *Journal of Money, Credit and Banking*, 47(8):1571–1598.
- Gelos, R. G., Sahay, R., and Sandleris, G. (2011). Sovereign borrowing by developing countries: What determines market access? *Journal of International Economics*, 83(2):243–254.
- Gennaioli, N., Martin, A., and Rossi, S. (2014). Sovereign Default, Domestic Banks, and Financial Institutions. *Journal of Finance*, 69(2):819–866.
- Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34.
- Gertler, M. and Kiyotaki, N. (2010). Financial Intermediation and Credit Policy in Business Cycle Analysis. In Woodford, B. M. F. . M., editor, *Handbook of Monetary Economics*, volume 3, pages 547–599. Elsevier, 1 edition.
- Gornemann, N. (2015). Sovereign Default, Private Investment, and Economic Growth. Working Paper.
- Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, Capacity Utilization, and the Real Business Cycle. *American Economic Review*, 78(3):402–17.
- Guembel, A. and Sussman, O. (2009). Sovereign Debt without Default Penalties. *Review of Economic Studies*, 76(4):1297–1320.

- Hatchondo, J. C. and Martinez, L. (2009). Long-duration bonds and sovereign defaults. *Journal of International Economics*, 79(1):117–125.
- Hatchondo, J. C., Martinez, L., and Sapriza, H. (2007). The economics of sovereign defaults. *Economic Quarterly*, 93(2):163–187.
- Iacoviello, M. (2005). House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle. *American Economic Review*, 95(3):739–764.
- Joo, H. (2014). Sovereign Default Risk, Fiscal Adjustment, and Debt Renegotiation. Working Paper.
- Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press. Google-Books-ID: 9Wxk_z9HskAC.
- Kletzer, K. M. (1994). Sovereign Immunity and International Lending. In Van der Ploeg, F., editor, *The Handbook of International Macroeconomics*. Blackwell, Oxford.
- Kumhof, M. (2004). Fiscal Crisis Resolution: Taxation versus Inflation. Working Paper 102004, Hong Kong Institute for Monetary Research.
- Mallucci, E. (2014). Sovereign Defaults, Wholesale Funding and Banking Crises. Working Paper.
- Mayer, M. (2011). Sovereign Credit Risk and Banking Crises. SSRN Scholarly Paper ID 1785893, Social Science Research Network, Rochester, NY.
- Mendoza, E. G. and Yue, V. Z. (2012). A General Equilibrium Model of Sovereign Default and Business Cycles. *The Quarterly Journal of Economics*, 127(2):889–946.
- Moody’s (2008). Sovereign Defaults and Interference: Perspectives on Government Risks. Technical report.
- Perez, D. (2015). Sovereign Debt, Domestic Banks and the Provision of Public Liquidity. Working Paper, Department of Economics, Stanford.
- Perry, G. and Servén, L. (2003). The anatomy of a multiple crisis : why was Argentina special and what can we learn from it? Policy Research Working Paper Series 3081, The World Bank.
- Reinhart, C. M. and Rogoff, K. S. (2008). This Time is Different: A Panoramic View of Eight Centuries of Financial Crises. Working Paper 13882, National Bureau of Economic Research.
- Reinhart, C. M. and Rogoff, K. S. (2011a). The Forgotten History of Domestic Debt*. *The Economic Journal*, 121(552):319–350.

- Reinhart, C. M. and Rogoff, K. S. (2011b). From Financial Crash to Debt Crisis. *American Economic Review*, 101(5):1676–1706.
- Richmond, C. and Dias, D. A. (2009). Duration of Capital Market Exclusion: An Empirical Investigation. SSRN Scholarly Paper ID 1027844, Social Science Research Network, Rochester, NY.
- Schmitt-Grohe, S. and Uribe, M. (2016). *Open Economy Macroeconomics*. Princeton University Press, Princeton.
- Sosa-Padilla, C. (2012). Sovereign Defaults and Banking Crises. MPRA Paper.
- Sturzenegger, F. and Zettelmeyer, J. (2007). *Debt defaults and lessons from a decade of crises*. MIT Press, Cambridge.
- Sturzenegger, F. and Zettelmeyer, J. (2008). Haircuts: Estimating investor losses in sovereign debt restructurings, 1998-2005. *Journal of International Money and Finance*, 27(5):780–805.
- Svirydzenka, K. (2016). Introducing a New Broad-based Index of Financial Development. IMF Working Paper 16/5, International Monetary Fund.
- Thaler, D. (2016). Reducing the computationally necessary state space by anticipating future choices. Working Paper.
- Yue, V. Z. (2010a). Sovereign default and debt renegotiation. *Journal of International Economics*, 80(2):176–187.

Appendices

Appendix A: The optimization problem of the government

Be X the vector of private sector variables $X = [C, L, D', K'_H, K'_B, B'_B, B'_X, K'_F, r, q, R', W, \lambda_1, \lambda_2, \lambda_3, \lambda_H]$. Then the government's problem, after substituting out some variables (HH value function, tomorrow's labor supply and dividend), expressing the foreign bond demand as a complementarity constraint and dropping a few redundant indices, is:

$$V_G(\Omega) = \max_{Rep, T, \bar{B}', X} \frac{C^{1-\gamma}}{1-\gamma} + \chi \frac{L^{1+\nu}}{1+\nu} + \beta E[V'_G]$$

s.t.

Government budget constraint

$$q\bar{B}' + T = \bar{B}Rep + G$$

$$B_F \geq 0$$

Bank FOCs

$$0 = rC^{-\gamma} - \beta E[V'_{Be}] - \lambda_1 E[C'^{-\gamma}] + \lambda_2 r$$

$$0 = -qC^{-\gamma} + \beta E[Rep'V'_{Be}] + \lambda_1 E[Rep'C'^{-\gamma}] \theta - \lambda_2 q + \lambda_3 q(1 - \psi)$$

$$0 = -C^{-\gamma} + \beta R' E[V'_{Be}] + \lambda_1 R' E[C'^{-\gamma}] \theta - \lambda_2 - \lambda_3 \psi$$

$$0 = \min \left\{ E[u_c(C)\theta(Rep' B'_B + R' K'_B)] - E[C'^{-\gamma} D'] , \lambda_1 \right\}$$

$$0 = \min \left\{ \eta(RK_B + RepB_B - D) - qB'_B - K'_B + rD' , \lambda_2 \right\}$$

$$qB'_B = (qB'_B + K'_B) \psi$$

Household FOCs

$$C^{-\gamma} = \beta/r E[C'^{-\gamma}]$$

$$C^{-\gamma} = \beta R/(1 + \xi) E[C'^{-\gamma}] + \lambda_h$$

$$0 = \min \left\{ \lambda_h , K'_H \right\}$$

$$C^{-\gamma} W = \chi L^\nu$$

Firm FOC

$$R' = E \left[\left(C'^{-\gamma} \omega' \right)^{\frac{1-\alpha}{\nu+\alpha}+1} \alpha K_F'^{\alpha-1} [(1-\alpha) K_F'^{\alpha} / \chi]^{\frac{1-\alpha}{\nu+\alpha}} \right] / E [C'^{\gamma}] + (1-\delta)$$

$$W = [\omega(1-\alpha) K_F^{\alpha} L^{-\alpha}]$$

Foreign lenders' bond demand

$$0 = \min \{q - E[Rep] \bar{q}, B_X\}$$

Market clearing

$$\bar{B} = B_B + B_X$$

$$K_F = K_B + K_H$$

Resource constraint

$$C + K_B' + K_H'(1 + \xi) + B_B' q = \bar{B}' q + \omega K_F^{\alpha} L^{1-\alpha} + (1-\delta) K_F - \bar{B} + B_B - G$$

Notice that in this problem, the government not only needs to form expectations over tomorrow's value function V_G' , but also over the marginal utility of consumption $u_c(C', L') = C'^{-\gamma}$, the marginal value of bank equity $V_{B_e}' = C'^{-\gamma} + \eta \lambda_2'$, the repayment choice Rep' and products thereof.

Appendix B: Computation

State space reduction:

The state space of the model described above is 7 dimensional $\Omega = [\omega, \bar{B}, K_H, D, B_B, K_B, R]$. To reduce the computational burden it is helpful to reduce its dimensionality. This can be done by anticipating the default decision. In the following I will explain this trick, which is an application of the method explained in Thaler (2016), in two steps.

First assume that for a given state Ω we know the default optimal decision. In that case, we can compute cash-at-hand (or pre-dividend equity) of the bank after debt repayment directly from the information contained in Ω :

$$E \equiv R K_B + Rep B_B - D$$

Cash-at-hand for the household, after debt repayment (but before receiving the proceeds of new debt issuance) is given by the difference of the total resources of the economy minus bank equity:

$$W = \omega (K_B + K_H)^\alpha L^{1-\alpha} + (1 - \delta) (K_B + K_H) - Rep(\bar{B} - B_B) - E$$

This variable contains the non-predetermined choice L though.¹⁰² Therefore, we need to keep track separately of

$$K_F = K_B + K_H$$

and

$$B_X = \bar{B} - B_B$$

Once we know the three variables $[E, B_X, K_F]$, plus the exogenous state ω , we have sufficient information to solve the governments problem for all variables other than Rep , that is we can find (T, \bar{B}', X) .¹⁰³ Call this alternative “state vector” $[\omega, \tilde{\Omega}] = [\omega, [E, K_F, B_X]]$. Assume we solve the governments problem across this alternate state vector.

But what about the repayment decision, which after all depends on all the 7 state variables in Ω ? This bring us to step two. Whenever we solve for (T, \bar{B}', X) given the state $[\omega, \tilde{\Omega}]$, we determine the values of all the endogenous elements in Ω . That means, we can determine the value of tomorrow’s endogenous alternative state *conditional on repayment*: $\tilde{\Omega}'|Rep'$. Furthermore assume that for each $(\bar{B}, K_H, D, B_B, K_B, R)$ there exists a threshold level of the exogenous shock $\bar{\omega}$, below which default is optimal and above which repayment is optimal. (This assumption is verified ex post.) Once we know this threshold value $\bar{\omega}$, we can determine tomorrow’s endogenous state as a function of the exogenous state: $\tilde{\Omega}'(\omega') = [E'(\omega'), B'_X, K'_F]$. This is all we need to know in order to compute the expectations over future variables, given we have approximated them across the state $[\omega, \tilde{\Omega}]$. Finally, we need to ensure that we picked the right $\bar{\omega}$. We know that at $\bar{\omega}$ the government must be indifferent between default and repayment, i.e. $V(\bar{\omega}, \tilde{\Omega}|Rep' = 1) = V(\bar{\omega}, \tilde{\Omega}|Rep' = 0)$. To find this threshold level, we therefore augment the above optimization problem by the variable $\bar{\omega}$ and the condition $V(\bar{\omega}, \tilde{\Omega}|Rep' = 1) = V(\bar{\omega}, \tilde{\Omega}|Rep' = 0)$. I call this trick “anticipation of future choices” because what we essentially do here is to explicitly anticipate one of the choices that the government has to make tomorrow and incorporate it into today’s problem - with the aim of being able to reduce the state space.

This means, that by complicating the optimization problem only slightly (by adding one equation and one variable), we are able to reduce the computationally necessary state space from 7D to 4D¹⁰⁴. This brings about a massive reduction in computation time.

¹⁰²If L is constant it is enough to know E and W . Furthermore, given separable utility, one could apply the trick that I apply to the repayment decision also to L . Then, even with variable labor, knowing E , W and ω is sufficient.

¹⁰³ To see this note that the state variables only appear in the above problem in these combinations.

¹⁰⁴As mentioned in the main text, a reduction to 5D is feasible without this trick, and a further reduction

Algorithm:

Apart from the way to deal with the state space and to evaluate integrals (expectations), the time iteration algorithm I use is a standard one loop algorithm, commonly used in the sovereign default literature.

1. Define a Cartesian grid over the 4D alternative state vector. Instead of using $[\omega, E, K_F, B_X]$ I rotate the grid so as to reduce the inclusion of regions of the state space into the grid, which are never visited along the equilibrium path. In particular I use $[\omega, \widetilde{W}, E/\widetilde{W}, K_F - (\varsigma_0 + \varsigma_1 W)]$ where $\widetilde{W} = (K_B + K_H)^\alpha + (1 - \delta)(K_B + K_H) - \text{Rep}(\bar{B} - B_B)$ and $\varsigma_0 + \varsigma_1 \widetilde{W}$ is the result of a regression of K_F on \widetilde{W} . I use (8,10,8,5) points.
2. Make an initial guesses for the functions $u_c(C)$, λ_2 and V_G across the points of this grid.
3. From these guesses and this grid, construct a cubic spline interpolant for each of the 3 functions. I use not-a-knot end conditions.
4. For each of the points on the grid, solve the optimization problem described in appendix A, augmented by the additional state $\bar{\omega}$ and the additional equation discussed above.
5. Check the difference between the previous guess and the solutions at the grid points obtained. If they are very similar stop. Else update the initial guess and return to point 3.

The whole code is written in MATLAB. To solve the nonlinear optimization problem at each of the grid points at step 4 of the algorithm, I use the solver KNITRO, which is able to solve smooth complementarity problems fast and reliably. To improve the performance of the solver, I supply analytical first derivatives, which are largely computed and coded automatically using MATLAB's symbolic toolbox. Furthermore the code is executed in parallel.

Despite the fact that the policy functions exhibit minor kinks but are approximated by smooth functions, the precision of the result is satisfactory: The algorithm converges successfully up to an average (across the grid) absolute change of the forward looking variables of 0.0001%. Across a very long simulation the mean absolute error of the forward looking variables (also known as Euler Error, see Judd (1998)) is around 0.05%.

to 3D is feasible by applying the same trick o the labor decision.

Chapter III

Reducing the computationally necessary state space by anticipating future choices

1 Introduction

Assume you want to find a recursive solution to a dynamic model with global solution methods. This dynamic model is given by the following equation system¹⁰⁵:

$$f_1(x_{t-1}, x_t, z_t, e_t, E_t[f_2(z_{t+1}, e_{t+1})]) = 0 \quad (1)$$

where x is a vector of n endogenous state variables, z is a vector of o forward-looking endogenous variables¹⁰⁶ and e is a vector of p exogenous state variables. A stationary equilibrium in this model is given by a set of policy functions $x(x_{t-1}, e_t) : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ and $z(x_{t-1}, e_t) : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^o$ that solve system (1).

To approximate this solution with global methods, for example with policy function iteration, parametrized expectations or collocation, one typically specifies a grid over the state variables and solves the equation system either sequentially or simultaneously at all grid points. The computational cost involved increases exponentially in the number of state variables, a problem known as the *curse of dimensionality*. Researchers therefore typically try to use the smallest computationally necessary state space possible. Let me briefly discuss two well known strategies to reduce the state space before I come to the method I propose, which is a generalization of the two more common approaches.

Manipulation 1: The most obvious trick is to eliminate redundant endogenous state variables. That is assume the system (1) can be rewritten as

$$\begin{aligned} f_3(w_{t-1}, x_t, z_t, e_t, E_t[f_2(z_{t+1}, e_{t+1})]) &= 0 \\ g_1(x_t) &= w_t \end{aligned} \quad (2)$$

¹⁰⁵Notice that the method is also applicable to constrained maximization problems and value function iteration. I restrict the exposition to an equation system merely because it eases exposition.

¹⁰⁶including purely contemporaneous variables

where $g_1(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $n > m$. I.e. $g_1(x_{t-1})$ is a function that summarizes the relevant state. f_3 is an arbitrary function. The solution of this manipulated system $(x, z)(w_{t-1}, e_t) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^n \times \mathbb{R}^o$ depends on less endogenous state variables and hence requires less computation time to solve for.

Manipulation 2: Another well known trick, by which the number of endogenous states can sometimes be reduced is to combine exogenous states and endogenous states in a new summary state variable. This is possible if we can rewrite (1) as

$$\begin{aligned} f_4(w_t, x_t, z_t, e_t, E_t[f_5(z_{t+1}, e_{t+1}, w_{t+1})]) &= 0 \\ g_2(x_t, e_{t+1}) &= w_{t+1} \end{aligned} \quad (3)$$

where $g_2(x, e) : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ with $n > m$. As before $g_2(x_t, e_{t+1})$ is a function that summarizes the computationally necessary state. f_4 and f_5 are arbitrary functions. Again, the solution of this manipulated system $(x, z)(w_t, e_t) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^n \times \mathbb{R}^o$ depends on less endogenous state variables and hence requires less computation time to solve for.

Manipulation 3: This trick can be generalized to allow for summary state variables which also include forward looking variables. Partition z_t into two parts, $z_{1,t}$ and $z_{2,t}$ of length o_1 and o_2 , i.e. $z_t = [z_{1,t}, z_{2,t}]$. Assume we can rewrite system (1) as

$$\begin{aligned} f_5(w_t, x_t, z_{2,t}, e_t, E_t[f_6(z_{1,t+1}, z_{2,t+1}, e_{t+1}, w_{t+1})]) &= 0 \\ g_3(x_t, z_{1,t+1}, e_{t+1}) &= w_{t+1} \\ h_1(x_t, E_t[h_2(z_{1,t+1}, e_{t+1})]) &= 0 \end{aligned} \quad (4)$$

where $g_3(x, z, e) : \mathbb{R}^n \times \mathbb{R}^{o_1} \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ with $n > m$. Again $g_3(x_t, z_{1,t+1}, e_{t+1})$ is a function that summarizes the computationally necessary state. Note that this function now also depends on forward-looking endogenous variables. f_5, f_6, h_1 and h_2 are arbitrary functions. The solution of this manipulated system is given by three functions $x(w_t, e_t) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, $z_2(w_t, e_t) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^{o_2}$ and $z_1(w_{t-1}, e_{t-1}) : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathcal{F}$ where \mathcal{F} is the function space of all functions $F : \mathbb{R}^p \rightarrow \mathbb{R}$. Again this solution depends on less endogenous state variables and hence is likely to require less computation time to solve for. Yet notice that while $(x, z_2)(w_t, e_t)$ maps from a hypercube of reals (the state space) to a hypercube of reals, $z_1(w_t, e_t)$ maps from the same hypercube of reals to a function space. This function essentially tells us tomorrows choice of z_1 conditional on tomorrows exogenous states.

Instead of explaining the idea any further at this abstract level, let us turn to an simple concrete example and consider the three proposed manipulations in turn.

2 Example: A neoclassical growth model

Consider a version of the neoclassical growth model with capital with full depreciation, labor and a storage technology. The household maximizes:

$$\begin{aligned} V &= \max_{c_t, s_t, k_t, l_t} u(c_t) - v(l_t) + \beta E(V') \\ \text{s.t.} \\ c_t + s_t + k_t &= l_t W_t + R_t k_{t-1} + R_s s_{t-1} \end{aligned}$$

where c , s , k , l , W , R , R_s and denote consumption, safe assets, capital, labor, the wage, the return on capital and the return on the safe asset respectively. t indicates the period in which a variable is determined.

The firm has a Cobb-Douglas production technology with variable capital utilization q and maximizes its profits:

$$\begin{aligned} \max_{l_t, k_{d,t}} y_t - W_t l_t - R_t k_{d,t} \\ \text{s.t.} \\ y_t &= A_t f(k_{d,t}, q_t) k_{d,t}^\alpha l_t^{1-\alpha} \end{aligned}$$

where TFP A_t follows a stochastic AR(1) process.¹⁰⁷

The first order conditions of the two agents problems together with the market clearing conditions ($k_{t-1} = k_{d,t}$) gives rise to a system of equilibrium conditions, which can further be simplified by substituting out R_t and W_t ¹⁰⁸

$$\begin{aligned} u_c(c_t) &= \beta E \left[A_t \left\{ f(k_t, q_{t+1}) \alpha k_t^{\alpha-1} + f_k(k_t, q_{t+1}) k_t^\alpha \right\} l_{t+1}^{1-\alpha} u_c(c_{t+1}) \right] \\ u_c(c_t) &= \beta R_s E [u_c(c_{t+1})] \\ c_t + s_t + k_t &= A_t f(k_{t-1}, q_t) k_{t-1}^\alpha l_t^{1-\alpha} + R_s s_{t-1} \\ v_l(l_t) &= u_c(c_t) A_t f(k_{t-1}, q_t) (1 - \alpha) k_{t-1}^\alpha l_t^{-\alpha} \\ f_z(k_{t-1}, q_t) &= 0 \end{aligned} \tag{5}$$

This equation system implicitly defines the equilibrium of this baseline economy. Notice how this system resembles the generic system (1): In this case the exogenous state is given by $e = A$ and the endogenous state by $x = [k, s]$. Furthermore the system features

¹⁰⁷This model is constructed to be as simple as possible yet to allow to illustrate all types of state space reduction. While this toy model serves as an example only, one can think of this model as a small open economy where the safe asset represents international lending.

¹⁰⁸One could reduce the system further by substituting c but i refrain from doing so here to keep the equations shorter.

3 forward looking variables $z = [l, c, q]$. A stationary equilibrium in this baseline economy is given by 5 policy functions $[c, s, l, k, q](k_{t-1}, s_{t-1}, A_t)$ that satisfy system (5).

I will now illustrate the 3 different ways to reduce the state space using simplified versions of this example.

Case 1 To illustrate the easiest way to reduce the state space consider first a simplified version of the baseline model where $A_t = 1$, $l_t = 1$ and $f = 1$, that is a deterministic version of the model without labor and without utilization. In that case we can reduce the number of states by summarizing s and k by cash-at-hand w .

$$w_t = g_1(s_t, k_t) = \alpha k_t^\alpha + R_s s_t$$

The state vector is now given by w_{t-1} instead of by (s_{t-1}, k_{t-1}) . This is a manipulation of the type described by equation system (2). This trick is well known and discussed in great analytical depth in the Kubler and Schmedders (2003), where they define the concept of a wealth recursive equilibrium for portfolio choice models.

Case 2 Next consider that we only fix labor $l_t = 1$ and utilization $f = 1$ but keep the TFP shock A_t . In that case we can again reduce the number of states by summarizing A , s and k by cash-at-hand w .

$$w_t = g_2(A_t s_{t-1}, k_{t-1}) = \alpha A_t k_{t-1}^\alpha + R_s s_{t-1}$$

The state vector is now given by (A_t, w_t) instead of by (A_t, s_{t-1}, k_{t-1}) . This is a manipulation of the type described by equation system (3).

Case 2b To build intuition for the third type of manipulation, let us first consider another more complicated case of the second type of manipulation. Consider the baseline model but without labor $l_t = 1$. In this case cash at-hand is given by

$$w_t = g_2'(A_t s_{t-1}, k_{t-1}, q_t) = A_t f(k_{t-1}, q_t) k_{t-1}^\alpha + R_s s_{t-1}$$

This expression contains one variable, q_t , which is not predetermined, hence the trick applied in case 2 does not apply directly. Yet, if the last equation of system (5) can be solved analytically for q_t , yielding $q_t = q(k_{t-1})$, we can use it to substitute q_t in the whole system.

$$\begin{aligned} u_c(c_t) &= \beta E \left[A_t \left\{ f(k_t, q(k_t)) \alpha k_t^{\alpha-1} + f_k(k_t, q(k_t)) k_t^\alpha \right\} 1 - \alpha u_c(c_{t+1}) \right] \\ u_c(c_t) &= \beta R_s E[u_c(c_{t+1})] \\ c_t + s_t + k_t &= A_t f(k_{t-1}, q(k_{t-1})) k_{t-1}^\alpha + R_s s_{t-1} \end{aligned} \tag{6}$$

Consider this reduced system to as the initial system (1). A solution to this system consists of 3 functions $[c, s, k](A_t, s_{t-1}, k_{t-1})$. Thanks to the elimination of q_t , cash-at-hand now does not depend on forward looking variables any more:

$$w_t = g_2''(A_t, s_{t-1}, k_{t-1}) = A_t f(k_{t-1}, q(k_{t-1})) k_{t-1}^\alpha + R_s s_{t-1}$$

We can use this definition to rewrite system (6) using a smaller state space (A_t, w_t) .

$$\begin{aligned} u_c(c_t) &= \beta E \left[A_t \left\{ f(k_t, q(k_t)) \alpha k_t^{\alpha-1} + f_k(k_t, q(k_t)) k_t^\alpha \right\} 1 - \alpha u_c(c_{t+1}) \right] \\ u_c(c_t) &= \beta R_s E [u_c(c_{t+1})] \end{aligned} \quad (7)$$

$$c_t + s_t + k_t = w_t$$

$$A_{t+1} f(k_t, q(k_t)) k_t^\alpha + R_s s_t = w_{t+1}$$

This manipulation results in a system of the generic form described by system (3). Yet notice: while in case 2 we just collected predetermined state variables, this time we had to first eliminate one forward looking variable by replacing q_{t+1} with the inverse of the relevant FOC. Hence we have reduced the set of endogenous variables by “anticipating a future choice” before we could reduce the state space. This more involved manipulation is used for example in Khan et al. (2014). They set up a model of heterogeneous firms, which employ labor and capital and finance the latter partially through borrowing and reduce the firms state space by one dimension by merging capital and borrowing to cash-a-hand, after substituting out the firms labor choice. To my k

Conceptually, the manipulation I propose and explain in case 3 is an extension of this manipulation for the case when reducing the set of variables is not analytically feasible.¹⁰⁹ To my knowledge this note is the first to explore this extension systematically, even though it appears not unlikely to have been used by researchers before.

Case 3 Finally let us consider a manipulation of the type described by equation system (4). For this purpose consider the model with labor and TFP but without utilization ($f = 1$). In this case cash at hand is given by

$$w_t = g_3(A_t s_{t-1}, k_{t-1}, l_t) = A_t k_{t-1}^\alpha l_t^{1-\alpha} + R_s s_{t-1} \quad (8)$$

¹⁰⁹In fact, adding q_t back as a variable and $q(k_t) = q_{t+1}$ as an additional equation to system (7) we arrive at an equation system that resembles system (4). This system can be directly derived from system (5) without changing the set of variables, just as the generic system (1) can be derived from the generic system (4).

Yet, using the more generally applicable but also more complicated computational approach I discuss in case 3 instead of the one discussed in case 2b is likely to be computationally inefficient. Of course the manipulation discussed in case 2b might be feasible, because $f_z(q_t, k_{t-1}) = 0$ might not be solvable analytically for q_t . In that case applying the method from case 3 becomes the only alternative. Note that this instead of adding I equations as in the example with labor, one would only need to add 1 equation, because f is independent of A .

This expression contains l_t , which is not predetermined. Hence we can not apply the same trick as in case 2 in a straightforward fashion. Furthermore we can not derive a closed form solution for l_t as a function of exogenous or predetermined variables from system (5), as we were able to for q_t in case 2b, because the labor supply equation contains the unknown function $c(k, s, A)$. Hence the trick from case 2b doesn't apply. Yet it is still possible to reduce the state space using cash-at-hand w_t by anticipating the choice l_t using an implicit function for l_t (instead of an explicit function as in case 2b for q_t).

We start from the initial system, which resembles the generic system (1)

$$\begin{aligned}
u_c(c_t) &= \beta E \left[A_t \left\{ \alpha k_t^{\alpha-1} \right\} l_{t+1}^{1-\alpha} u_c(c_{t+1}) \right] \\
u_c(c_t) &= \beta R_s E [u_c(c_{t+1})] \\
c_t + s_t + k_t &= A_t k_{t-1}^\alpha l_t^{1-\alpha} + R_s s_{t-1} \\
v_l(l_t) &= u_c(c_t) A_t (1 - \alpha) k_{t-1}^\alpha l_t^{-\alpha}
\end{aligned} \tag{9}$$

A solution to this system is a function $[c, s, l, k] (k_{t-1}, s_{t-1}, A_t) : \mathbb{R}^3 \rightarrow \mathbb{R}^4$. We manipulate this system by shifting the labor equation one period forward and using the definition of cash at hand (8):

$$\begin{aligned}
u_c(c_t) &= \beta E \left[A_t \left\{ \alpha k_t^{\alpha-1} \right\} l_{t+1}^{1-\alpha} u_c(c_{t+1}) \right] \\
u_c(c_t) &= \beta R_s E [u_c(c_{t+1})] \\
c_t + s_t + k_t &= w_t \\
v_l(l_{t+1}) &= u_c(c_{t+1}) A_{t+1} (1 - \alpha) k_t^\alpha l_{t+1}^{-\alpha} \\
w_{t+1} &= A_{t+1} k_t^\alpha l_{t+1}^{1-\alpha} + R_s s_t
\end{aligned} \tag{10}$$

This system now resembles the generic system (4) and depends on no state variables other than w_t and A_t . A solution to this system consists of 2 elements: First, a policy function $[c, s, k] (w_t, A_t) : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ that describes the contemporaneous choice of c , s and k . Second, a function $l(w_t, A_t) : \mathbb{R}^2 \rightarrow \mathcal{F}$ where \mathcal{F} is the function space of all functions $F : \mathbb{R} \rightarrow \mathbb{R}$. This function l describes, for a given state of the world today, the labor choice that pertains in the next period conditional on the realization of the exogenous state next period.

An intuitive way to think about this representation of the original problem is to split each period in 2 sub-periods: in sub-period 1 agents produce and in sub-period 2 they spend their income. Instead of solving for sub-period 1 and 2 of period t simultaneously we solve for sub-period 2 of period t and sub-period 1 of period $t+1$ simultaneously.¹¹⁰

¹¹⁰Unlike the solution to the original problem, the solution to the modified problem does not provide us with a policy function of current labor as a function of the current state. This is natural since we have defined the state to be “post labor choice”. Furthermore this does not reduce the usefulness of the solution. Simulations including the path for labor for example are still possible: Starting at a given state

The unusual part about this solution is the function l . Next I will explain how this solution can be computed using policy function iteration.¹¹¹ But first let me recapitulate how to find the solution to standard representation of the problem with the bigger state space (system (9)).

With policy function iteration we typically approximate the forward looking variables by some approximating function (e.g. interpolation) over the state space. In our case that means we might replace c_{t+1} by the approximating function $\tilde{c}(k_{t-1}, s_{t-1}, A_t)$ and l_{t+1} by $\tilde{l}(k_{t-1}, s_{t-1}, A_t)$. Furthermore we usually approximate expectations with quadrature methods. That means we replace the expectations of variable x , $E[x_{t+1}(A_{t+1})]$, by $\sum_{i=1}^I \pi_i x(A_i) = \sum_{i=1}^I \pi_i x_i$ where I is the number of grid points for the exogenous state A and π_i are the corresponding weights. Let us write the equation system (9) using these approximations explicitly:

$$\begin{aligned} u_c(c_t) &= \beta \sum_{i=1}^I \left[\pi_i \left\{ A_i \alpha k_t^{\alpha-1} \tilde{l}(k_{t-1}, s_{t-1}, A_t)^{1-\alpha} \right\} k_t u_c(\tilde{c}(k_t, s_t, A_{t+1})) \right] \\ u_c(c_t) &= \beta R_s \sum_{i=1}^I [\pi_i u_c(\tilde{c}(k_t, s_t, A_{t+1}))] \\ c_t + s_t + k_t &= A_t k_{t-1}^\alpha l_t^{1-\alpha} + R_s s_{t-1} \\ v_l(l_t) &= u_c(c_t) \left\{ A_t (1 - \alpha) k_{t-1}^\alpha l_t^{-\alpha} \right\} \end{aligned}$$

Traditional policy function iteration would now start with some guesses for the functions \tilde{c} and \tilde{l} and solve this system for each grid point on a grid for $[k_{t-1}, s_{t-1}, A_t]$. Then the guesses would be updated. One would iterate over the last two steps till convergence.

To solve the alternative representation of the problem given by system (10) we plug in cash-at-hand w_t and use it as a state instead of $[k_{t-1}, s_{t-1}]$. Furthermore, instead of defining an approximating function \tilde{l} we “anticipate” the labor choice l_{t+1} by adding the labor optimality condition for tomorrow (and dropping the labor optimality condition for today). Since we need to know the labor supply for each state tomorrow we have to add I such equations. Notice that these I equations together approximate the function to

defined by w_0 and A_0 the labor policy functions gives us today's choices of k_0 , c_0 and s_0 and tomorrow's labor conditional on all possible values of A_1 . Then we iterate forward, draw a new A_1 , use this A_1 to pick the relevant l_1 and combine A_1 , l_1 , s_0 and k_0 to get w_1 .

¹¹¹See below for the applicability of this state space reduction method to other global solution methods.

which $l(w_t, A_t)$ maps.

$$\begin{aligned}
u_c(c_t) &= \beta \sum_{i=1}^I \left[\pi_i \left\{ A_i \alpha k_t^{\alpha-1} l_i^{1-\alpha} \right\} k_t u_c(\tilde{c}(w_i, A_i)) \right] \\
u_c(c_t) &= \beta R_s \sum_{i=1}^I [\pi_i u_c(\tilde{c}(w_i, A_i))] \\
c_t + s_t + k_t &= w_t \\
v_l(l_1) &= u_c(\tilde{c}(w_1, A_1)) \left\{ A_1 (1 - \alpha) k_t^\alpha l_1^{-\alpha} \right\} \\
&\vdots \\
v_l(l_I) &= u_c(\tilde{c}(w_I, A_I)) \left\{ A_I (1 - \alpha) k_t^\alpha l_I^{-\alpha} \right\}
\end{aligned} \tag{11}$$

where $w_i = A_i k_t^\alpha l_i^{1-\alpha} + R_s s_t$. With this system we proceed as before: We start with a guess for \tilde{c} , solve the system at each grid point, update \tilde{c} and iterate to convergence.

This new equation system has $3 + I$ equations and $3 + I$ variables $(k_t, s_t, c_t, l_1, \dots, l_I)$. Due to the larger number of variables, solving this system at each grid point may be somewhat more time consuming, say X times more time consuming than solving the original equation system with 4 equations and 4 variables. Yet we have to solve the new system for each point on a 2D grid, while the original system had to be solved for each point on a 3D grid. That means that each time iteration will take roughly J times longer with the original method, if J is the number of grid points for the endogenous variables. With a efficient solver (e.g. using analytical derivatives) and with a fine enough grid for the endogenous variables but a coarse enough grid for the exogenous variables, it is very likely that $X \ll J$, hence that the modified problem is much faster to solve.

3 Applicability of the method

Let me conclude with a few remarks about the scope of applicability of this method.

Types of models

First, notice that this method can of course be use to reduce the number of states by even more than one as in the above example. For example consider a two country version of our model. In that case we could reduce the state space from 4 endogenous variables to 2 endogenous variables. Or consider a version of the model with human capital. In that case we could reduce the state space from 3 endogenous variables to 1 endogenous variables.

While these models are useful as an example, the method is applicable to more complicated and more interesting macroeconomic models as well. In a paper on sovereign default with 2 agents and 5 assets Thaler (2016) I apply this method to anticipate the default threshold, which allows me to reduce the state space from 7 to 4 dimensions.

While not limited to this class of models, these examples show that a natural class of models, for which this method is useful are models featuring portfolio selection.

Yet this method is not universally applicable. In particular the method requires that one can summarize the states in a convenient form such that the resulting system features the endogenous variables that enter the summary variable only with time indexes higher than the time index with which they enter the summary variable. In the example above the endogenous variables entering w_t are l_t , s_{t-1} and k_{t-1} . Note that none of these variables appear (with that index or a lower index) in system (11). It is easy to find examples that violate this assumption. For example assume a non-separable utility function. Then l_t would appear on the LHS of the two Euler equations in system (11) in $u_c(c_t, l_t)$. Knowing cash-at-hand would then not be sufficient any more to describe the relevant state of the economy. One could of course add l as a state variable, but this would only make sense if the summary variable eliminated more than 1 state (as in the extension with human capital). Another example that makes the method inapplicable are capital adjustment cost $\phi(k_t/k_{t-1})$. They would show up on the LHS of the budget constraint. It would hence not be possible anymore to eliminate all references to k_{t-1} by plugging in w_t .

Types of solution methods and computational costs and benefits

Policy or Value function iteration We have seen how this method can be applied in a policy function iteration context. It is straight forward to apply the same method to value function iteration problems.

For these algorithms the additional computational cost of the reduction of the state space increases in the number of exogenous variables and the number of grid points for each exogenous variable. The benefits of the reduced state space increase with the number of endogenous variables that can be summarized and the number of grid points for each endogenous variable.

For the two country version of our example model the solution algorithm would speed up due to the reduction of the number of endogenous states from 4 to 2. Yet this would come at the cost of I ($2I$) extra equations assuming global (regional) TFP shocks. In the default model Thaler (2016) the state space reduction from 7 to 4 dimensions comes at the cost of only 1 additional equation.

Collocation or weighted residuals The same method can be applied in a straightforward way when using collocation or weighted residuals methods. The computational costs and benefits should be similar as above.

Parametrized expectations Finally, the method can also be adapted for parametrized expectations algorithms. Given certain parametrized expectations $E_1(w_t, A_t)$ one first solves:

$$\begin{aligned}
u_c(c_t) &= \beta E_1(w_t, A_t) \\
u_c(c_t) &= \beta R_s E_2(w_t, A_t) \\
c_t + s_t + k_t &= w_t \\
v_l(l_t) &= u_c(c_t) A_t (1 - \alpha) k_{t-1}^\alpha l_t^{-\alpha}
\end{aligned} \tag{12}$$

This yields policy functions $(c, l, k, s)(w, A)$. Next one simulates the economy. Finally one updates the parametrized expectations $E_1(w_t, A_t)$ and restarts with step 1. In the simulation step one then has to solve after each period for the next period state by solving the equation

$$A_t k_{t-1}^\alpha l(w_t, A_t)^{1-\alpha} + R_s s_{t-1} = w_t$$

Whether the method still speeds up the process now depends on the cost of solving this equation and the length of the simulation.

References

- Khan, A., Senga, T., and Thomas, J. (2014). Default risk and aggregate fluctuations in an economy with production heterogeneity. The Ohio State University Working Paper.
- Kubler, F. and Schmedders, K. (2003). Stationary equilibria in asset-pricing models with incomplete markets and collateral. *Econometrica*, 71(6):1767–1793.
- Thaler, D. (2016). Austerity to save the banks? A quantitative model of sovereign default with endogenous default costs and a financial sector. Working Paper.